UNIVERSITY OF SOUTHAMPTON

FACULTY OF BUSINESS, LAW AND ART

School of Law

Towards A New Approach To The Legal Definition Of Personal Data And A Jurisdictional Model Of Data Protection Law: Surpassing The Requirement For An Assessment Of Identifiability From Data With An Effects-Based Approach

by

Alison Mary Knight

Thesis for the degree of PhD

November 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF BUSINESS, LAW AND ART

Law

Thesis for the degree of Doctor of Philosophy

TOWARDS A NEW APPROACH TO THE LEGAL DEFINITION OF PERSONAL DATA AND A JURISDICTIONAL MODEL OF DATA PROTECTION LAW: SURPASSING THE REQUIREMENT FOR AN ASSESSMENT OF IDENTIFIABILITY FROM DATA WITH AN EFFECTS-BASED APPROACH

Alison Mary Knight

Means of identification are growing rapidly with new digital and online tracking capabilities, and other emerging technologies of personal verification and authentication. Today, we emit a wide spectrum of direct - but also indirect – identifiers which, in recorded form, either alone or in combinations, can lead to us being known in different ways. Indeed, in a world that is becoming increasingly hyper-connected and digitally-surveyed, internet-connected devices leak a constant stream of data with which individual users (and proximate others in the data-collection vicinities) can be associated. Furthermore, the more data about a person that are collected, the easier it can become for further information to be inferred about them, linked to them, as well as for them to be singled out from others. Along with the development of data analytical techniques, this has led to recognition that scientific and technological advances are making it increasingly difficult to de-identify personal data *and* guarantee that the individuals to whom such data relates may not be later re-identified from it.

The legal, privacy, and regulatory challenges that flow from these facts have led to the present research. They point to the fact that there is confusion about the legal definition of personal data reliant upon the concept of identification capabilities from information, in interpretation and practical application (in determining what data comes within its scope and data protection obligations apply), when confronted with new technological realities. More specifically, such challenges point to the critical importance of reconsidering the requirement of being identified or identifiable from data as a key trigger for the application of data protection rules.

This research explores the significance of the legal requirement that an individual must be identified or identifiable from data as tantamount to the primary factor in determining whether it is personal data or not and subject to EU data protection law, both now and under incoming legal reform (the EU General Data Protection Regulation, 'GDPR') to take effect from 25 May 2018. The research findings enable an assessment to be made of the 'fitness-for-purpose' of an identificatory approach to personal data as providing a meaningful boundary to data protection regulation. The yard-stick of evaluation is achievement of the twin aims of EU data protection legislation: upholding the fundamental rights of data subjects (in particular, their right to privacy) in connection with the processing of data about them to a high level of equivalence in EU national laws, while also facilitating intra-EU / transborder-with-EU flows of such data.

Such an approach is compared against the contours of a new theoretical approach to personal data classification revolving around an analysis of the likely negative effects (risk of harm) to a data subject flowing from data processing activities intended to be applied to information. This comparison is then elaborated on via interrogation of the two approaches, and models founded upon such approaches (existing and prospective), in relation to determining when personal data may – and may no longer be - deemed personal under data protection law.

This thesis argues that reconceptualising the definition of personal data as dependent on an effects-based assessment, not an identificatory one, could lay the foundations for a more conceptually-coherent, jurisdictional methodology for determining when and, ultimately, what data protection rules should apply in context. Such a methodology would require a risk-management exercise to be carried out by those planning to process data relating to persons, involving the quantification of likely harm that may flow from particular data processing activities in context.

Consequently, a proposition is advanced to implement effect-based exemptions under data protection law, which takes insight from an effects-based framework already evolved under modernised EU competition law. In particular, consideration is given to the use of certain legal 'safe harbour' instruments - specifically, block exemption regulations providing comfort from enforcement action - as inspiration for moving to a more coherent, flexible and proportionate regulatory system able to take in to account collective interests. The resulting expansion of the data protection regulatory toolkit could incentivise and enhance confidence in compliance, as well as help encourage data sharing to promote innovation and demonstrable public benefits. Such a development would marry well with the incoming principles of accountability, and impact assessments, as the regulatory 'lynchpins' against which those who intend to process data relating to living individuals must abide by in the future as analytic practices and technological applications become more complex.

DECLARATION OF AUTHORSHIP

I, ALISON MARY KNIGHT

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

'TOWARDS A NEW APPROACH TO THE LEGAL DEFINITION OF PERSONAL DATA AND A JURISDICTIONAL MODEL OF DATA PROTECTION LAW: SURPASSING THE REQUIREMENT FOR AN ASSESSMENT OF IDENTIFIABILITY FROM DATA WITH AN EFFECTS-BASED APPROACH'

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed: .....

Date: .....

Table of Contents

Acknowledgements

Thank you to Professors Stephen (Saxby), Sophie (Stalla-Bourdillon), Mark (Weal), Kieron (O'Hara), Christopher (Millard), Sarah (Stevenage), Raj (Muttukrishnan), and Clare (Sullivan) for inspiring and helping me along the path to academia over the last 5 years and going forward.

Thank you to my friends in Web Science for welcoming me into a wonderful, interdisplincary academic community, and for my friends in Legal Services for keeping me grounded in practice.

Thank you to my colleagues on:

- The Super-Identity Project (2012-2015, EPSRC Grant EP/J004995/1), "*provid[ing] a step-change in the way that we think about identity and identification, and in the value that it might hold for the real world*"; and,

- The Data Pitch Project (2017-present, H2020-2.1.1 Grant 732506), "*help[ing] organisations and businesses tap into the potential of their data".*

Thank you to the members of my immediate and extended Knight family (including members of the the Macleans, the Slades, the Logans, and the Rostens) for always being there for me, not forgetting the constant love of the two 'Ls'!

Dedicated to Martin, without whom I wouldn't be 'me'.

Abbreviations

Charter - Charter of Fundamental Rights of the EU (2000/C 364/01)

CJEU - Court of Justice of the EU

Convention 108 - Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data 1981

DPA – UK Data Protection Act 1998

DPAs – EU Member States' Data Protection Authorities

DPD – Data Protection Directive (officially Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data)

DPIA(s) – Data protection impact assessment(s)

ECHR – European Convention on Human Rights

ECtHR – European Court of Human Rights

EU – European Union

FIPs – Fair information principles

FOIA – UK Freedom of Information Act 2000

FOIASA - Freedom of Information (Scottish) Act 2002

GDPR – General Data Protection Regulation (officially Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC)

ICO – UK Information Commissioner's Office

IoT – Internet of Things

MS(s) – EU Member State(s)

PIA(s) – Privacy impact assessments

WP – EU Article 29 (Data Protection) Working Party

TGN – Technical Guidance Note

TFEU – Treaty on the Functioning of the EU

UK – United Kingdom

US – United States

Foreword

"MR. CALO: Quickly, I think there is a big difference between anonymized and aggregate, first of all. I just -- it's like I don't really care if -- I mean, imagine a consumer who says, I hate advertising so much that I don't want any of my data to go towards those advertisers and so that's a sticking point for them, right. So apart from that rare person, anonymized, does that really matter? Does that really matter if they know who you are? I never sort of -- I mean, I understand the importance of anonymization, of course. And I've read Paul Ohm's excellent work like everyone else, but at the end of the day, like let's say that after you have a 12 mile run, that's sort of one of the scenarios. You have a 12 mile run and you are on this app and what it does it is tells Snickers that you just completed a 12 mile run. And Snickers then is able to send you a text to your phone saying here's a coupon for Snickers, here's the closest place to get Snickers, right? And here you have run, you're so good, you've run and burned off all those calories and then all of the sudden, oh, you're susceptible. And this is when you get the Snickers ad, right? I mean -- think about the New York Times –

MS. MITHAL: But is that really anonymized or aggregate?

MR. CALO: Well, that's just what I'm saying. So does it matter if they know who I am? It could be utterly anonymized. It could just be device 1234 went for a 12 mile run, do you know what I mean?

MS. MITHAL: Yeah.

MR. CALO: It doesn't matter who it is. And so for me, those are different threat scenarios.

MS. MITHAL: Right, right. So one scenario is, they don't know that you are Ryan Calo, but they know that you are device 1234. Another scenario is Snickers gets the information of a 1,000 runners and says here's where we need to place our billboards. So those are two separate scenarios. You know, I think that the example, if it is anonymous and de-identified, sort of gets to a larger question that we've got to think through as sort of, what's the harm? I mean, we might not like the scenario, you know, of running a marathon and then getting a Snickers bar, but in the overall scheme of things, is that really harmful as a consequential -- I mean, we've had a lot of discussion today about medical information or we protect financial information or kids' information. I think we need to think through some of the consequences …"

US Federal Trade Commission (FTC) 'Internet of Things privacy and security in a connected world' - 2013 workshop transcript

---

"The value of identity of course is that so often with it comes purpose"

Richard. R. Grant

Chapter 1 - Introduction

This chapter describes the rationale for this thesis and sets the stage for the analysis undertaken in successive chapters. The thesis theme is the concept of personal data under European Union (EU) data protection law, and specifically its requirement that someone be identified or identifiable from such data, and considerations around a new approach. Said simply, is this identificatory requirement still fit for purpose in keeping pace with challenges associated with modern technology, or might a replacement be more effective?

The Chapter divides into two parts for reader ease. In Part I, there are two sections:

- **Section 1.1 –** sets the scene in describing how our relationship with data has changed in recent times, and how this fact raises certain legal challenges.
- **Section 1.2** – outlines the factual and theoretical challenges that frame the resulting problem statement. It also introduces the value of conceptualising identifiability from data as a spectrum of identification risk.

In Part II, there are nine sections:

- **Sections 1.3-1.5 –** describe the research rationale, purpose, and objectives, alongside the thesis research questions.
- **Section 1.6** – outlines the novelty of the thesis, and its contribution aims.
- **Section 1.7 –** provides an overview of the thesis structure.
- **Sections 1.8-1.9 –** describe the research methodology while acknowledging certain limitations on research scope.
- **Section 1.10 –** introduces a case study for illustrating this thesis' main ideas as it develops.
- **Section 1.11 -** draws chapter conclusions.

Part I – The Thesis Problem

1.1 Scene setting - our changing relationship with data and how it begs certain important legal questions

Today, much more data[1] are generated, stored, analysed, and shared about us and with us than ever before, facilitated over the last three decades by the World Wide Web's invention, increased processing power, and new ways to communicate (e.g. via mobile phones, and using social media). As a society, we are recording and communicating more and more data, which, in one way or other, may be associated with individuals.

This growth of digitisation and so-called 'data-fication' (enabling easier analysis of data across different formats/mediums) has ramifications for the scale of data that are now available to process and the types of automated processing activities. Many different data sources are spawned through ever-increasing publicly accessible data and more opportunities for data connectivity. Along with the development of stronger analytics, these facts also hint at the relative ease with which it is possible now to extract detailed information about people and re-analyse it for purposes different from those originally pursued upon its creation. Furthermore, the low cost of storage, combined with the ubiquitous availability of cloud computing services, means that data can be stored indefinitely as long as the relevant hardware and software enabling access still exists. Expressed differently, there are no significant technological reasons why data created today cannot be accessed and further processed in the future in the same quality as it was preserved. Large scale and quick data dissemination is also possible now, giving rise to a myriad of unknown potential secondary re-users of disseminated information to whom it may be disclosed.

These technological realities - and the increasing societal importance of information communications generally – therefore bring new challenges, in particular regarding the protection provided to individuals in relation to the processing of information about them, the potential extraction of detailed sensitive information from it, and a lack of transparency over what happens to it after its initial collection processing purpose is fulfilled.

[1] **In this thesis, the terms 'data' and 'information' are used interchangeably to refer to that which can be processed digitally by means of automated technologies. For ease, the term 'data' is also used below as a plural as well as singular noun (i.e. "the data are", as well as "the data is").**

1.1.1 The concept of personal data in law

Data protection and privacy laws typically step in to regulate certain processing activities involving data about people, notably to ensure protection in processing situations considered worthy of legal intervention by virtue of a piece of information's relationship to a particular individual. Data protection law, in particular, is also about another central relationship - between the subject of personal data and its controller; and, specifically, the interests of the former in having restrictions placed upon data processed about them by the latter (and those who process the data on their behalf).

Personal data is a key concept for protecting individuals under EU data protection law as its processing triggers the application of data protection legal principles (see Chapter 2) and associated requirements to protect the interests of the individuals about whom personal data is being processed by ensuring that it is handed fairly and lawfully.[2] Said otherwise, whenever personal data are processed data protection rules are switched on and, in that context, the application of such rules may be considered binary (either they apply because such trigger conditions are met, or they do not). Whereas, organisations that process personal data may be subject to regulatory action if they do so without fulfilling their legal responsibilities outlined under such rules.

The purpose of this thesis is to delve into the legal definition of personal data's conceptual 'mapping' – in terms of the exact nature of the links that must be found between an individual and specific information about them – and such that lawmakers have considered its processing a trigger factor for data protection law to apply. It seeks to do this via the articulation of two approaches underpinning this concept: one found in existing law (an 'identification-based' approach), and one postulated (an 'effects-based' approach). The exploration of each approach in turn should assist in understanding when the sufficiency of the strength of the link between data and a person requires that the processing of the former be subject to EU legal regulation. It also explores how interpretations of the legal concept of personal data have adapted over recent years to cope with emerging technologies, resulting in a blurring of the 'line' denoting the scope of legal protection afforded in connection with data linked to individuals. Moreover, it discusses how that notional line might be redrawn in terms of the scope of legal protection afforded in connection with processing data linked to individuals.

[2] Concepts similar to personal data are also prevalent outside the EU, e.g. the federal Australian Privacy Act 1988 refers to 'personal information', as does the federal Canadian Personal Information Protection and Electronic Documents Act 2000, albeit with different legislative tests for determining its existence. Comparisons can also be drawn with the concept of 'personally identifiable information' (hereafter, PII) triggering the application of certain federal (and some state-level) US privacy laws.

1.2.1 Identification and the legal concept of personal data

The deliberate choice by the EU legislator to refer to identificatory-capacity in the definition of personal data within data protection law (see Chapter 2)[3] should be viewed against the backdrop of the fact that personal data (by virtue of the term 'personal') is necessarily about a person, but not just any person – a particular person, in some way who must be 'pin-pointable' from the data. Chapter 2 develops the argument that – while other sub-requirements of the definition also have significant impact upon classification - the key phrase to understand in the definition of personal data under EU data protection law is "identified or identifiable". Information relating to persons is not necessarily personal data legally unless it is necessarily identifying of them. As will be seen, it is also this (identificatory) legal requirement/element that has beguiled regulators and scholars the most (as compared with the other definitional element-requirements) in terms of its meaning and impact on practical decision-making as to whether data should be classified as personal or not.

This thesis contends that this area is ripe for further research exploration, in particular around when information with the mere potential to be identifiable (personally revelatory in particular ways) of its subject is personal data. Conversely, how much destruction of this link (between a person and a particular piece of data) is sufficient to merit its processing no longer to be subject to the application of data protection law (and the underlying interests the law is intended to protect)? In turn, this requires clarifying what exactly is meant by "identified or identifiable" in relevant legislation and by those who have formally interpreted (and re-interpreted) this element in light of emerging technology capabilities. It also begs the question whether the identificatory requirement/element remains a valid criterion for determining the personal data definition in law. Can we make a case for a rigorous re-consideration of the conceptual foundations of this element and its meaning through revisiting the interests intended for protection via its role?

1.2 Problem statement

This section describes the factual challenges prompting this research as a backdrop to explaining why it was decided to evaluate the fitness-for-purpose of the identificatory requirement in this thesis.

[3] Under the EU Data Protection Directive (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), personal data is defined as "*any information relating to an identified or identifiable natural person ('data subject')*" (Article 2(a)).

1.2.1 The notion of identification becomes increasingly ambiguous when confronted with recent technological and societal trends

In the 20th century, what was mean by identifying-information was relatively clear. Formal methods for identifying members of the public were typically paper-based, appearance-centred, and transparent (e.g. via the checking of passports). Modern technology confuses this issue by transforming identification capabilities involving data. Specifically, the notion of identification from information has been shaped and co-evolved alongside socio-technical developments, as illustrated by three recent trends described in the following sub-headings. [4]

1.2.1.1 Personal device proliferation and the phenomenon of hyper-connectivity

Online and offline, each of us produces data when we interact with information technologies in, what is now, the process of carrying out ordinary activities via our personal devices (e.g. phones, tablets, and personal computers either standalone or in our cars and household objects). Through interacting constantly with our devices, we generate data collected by embedded software applications typically designed to feedback information about us so that knowledge can be commercialised. Whether we know it or not, we are constantly interacting with organisations trying to find out more about us.

Indeed, while nowadays people are voluntarily posting huge amounts of data about themselves online, many sources of information revelatory of us now transcend those that we either volunteer, or knowingly impart, to encompass non-obvious data from observing us unawares.[5] Furthermore, this happens in great quantities and frequencies, revealing information about us dynamically and in complex ways because of the widespread uptake and use of devices connected to digital networks.[6] Linked to these facts, we are fast becoming an always-on society, and increasingly "hyper-connected" individuals[7] within it, typically relying on a variety of different, Internet-connected devices to navigate our daily lives.

[4] Arguments extrapolated from Knight & Saxby (2014, pp.620-621).
[5] That is, it encompasses data captured often without the active cooperation, or full knowledge (in terms of what and why data is being collected exactly), of the persons to whom it relates.
[6] Examples include data generated indirectly about us from retail transactions, items on phone bills, smart meter energy usage patterns, and event data recorders in cars transmitting information back to businesses such as insurers.
[7] UK Government, 2013, Foresight Project Final Project report. Future Identities: Changing identities in the UK – the next 10 years, p.51.

1.2.1.2 The disappearance of untraceability - tracking technologies and new possibilities for covert persistent surveillance

Recording of online activities and mobile device locations increases possibilities to track people's movements and activities in the physical world. More specifically, we produce trails of 'metadata' (broadly, data about data)[8] whenever we interact with devices and, whether we realise it or not, such trails can often be used to monitor us over time and analyse our behaviour.

Consider automatic number plate recognition: a digital image of a licence plate number captured by a traffic camera uploaded to a real-time traffic monitoring system. It can be used to track the whereabouts of both the vehicle and the driver. Another offline example is radio-frequency identification (RFID) tags embedded into objects, such as consumer items. While such tags are directly linked to objects, rather than people, data emitted can be used to gather information about the locations of both. Location tracking may also occur using the built-in global-positioning system (GPS) functions of mobile devices, and media access control (MAC) addresses,[9] enabled by cellular networks (intercepting phone towers). Intercepting Google Map queries on mobile devices, as well as other software, may also reveal location data. Indeed, user information (not just about geo-location, but also details related to usage-uptime) may be gleaned from any type of device that processes data associated with human activities, especially those connectable to communication networks (such as via the Internet and the Web that operates over it).

Obvious online examples linked to devices and their users are data logs associated with Internet Protocol (IP) addresses[10] needed by any device connecting to the Internet, and cookies[11] associated with web browsers, both of which may contain unique identifiers. While the issue of 'who exactly'

---

[8] See, for example, the following description of metadata (in US Executive Office of the President and Podesta, J., 2014. Big data: Seizing opportunities, preserving, pp.34-35): *"a term describing the character of the data itself".* Examples from the field of communications include the numbers originating and terminating phone calls, browser history data, geographical coordinates, and time stamps. While they may be considered less revealing than the content of such communications, the potential sensitivity associated with such metadata is increasingly being recognised. This led the aforementioned report to conclude, "*those who argue that metadata today raises more sensitivities than in the past make a sufficiently compelling case to motivate review of policy on the matter"*, leading to consideration of *"the extent to which data and information should receive legal or other protections on the basis of how much it reveals about individuals".*

[9] A MAC address of a computer is its unique hardware number.

[10] To connect to the Internet, each device requires an IP address - a unique address (four to 12-digit number) – enabling it to request and receive content from websites and that can be recorded by the website operator. IP addresses are assigned by internet service providers and managers of local area networks. This can be done temporarily, so that they change every time a user logs in ('dynamic IP addresses'), or permanently to a user's device ('static IP addresses'). Chapter 3 discusses this further.

[11] Cookies are text files (typically strings of numbers) that can be placed on users' web browsers to allow website companies to track their online activities. First-party cookies are placed by the website that the user is visiting. Third-party cookies are typically used by advertisers and other aggregators (such as search engines and social networking sites), e.g., DoubleClick tracked by Google. For instance, a website might set a cookie on a user's browser with a key (its name) and value (the unique identifier assigned to a computer), which can be accessed by the third party on it server to track how a particular user visits different websites, and combine this information with other data (such as search queries and results recorded by a search engine) to make inferences about users.

is using a device may not be readily apparent from either alone, this may become more readily ascertainable when combined with other data (e.g., data held by a specific user's internet service provider, or search queries, linked to an IP address).[12]

Information revealed via devices and digital networks, therefore, potentially provides "*important insights into who we are, what we do, whom we do it with, and when and where we do it*".[13] Moreover, while fragmentary data about our activities may not provide much revelatory value on their own, in combination they can often reveal personal details in unanticipated ways.[14] Different elements of our digital 'footprints' can feed into detailed 'pictures' about each of our movements/behaviours, which become richer when further combined and cross-referenced to form a strong link between data and a person. For example, software can now enable the aggregation of information gathered from social networking sites, clickstream data, geo-location data, financial transactions, and IP network logs to build a detailed mosaic of an individual's movements, their essential characteristics, and interests. The search for this knowledge is also the foundation of the emergence of a revenue model for online companies relying on tracking online activities. This coincides with another recent trend.

1.2.3 Big data profiling patterns can be used to infer new knowledge about individuals

The rising phenomenon of 'big data'[15] - broadly, the evolution of large-scale analytics applied to increasing quantities of data with new capabilities on scale, speed, and data pattern recognition –

---

[12] See, Segrist (2014, p.545) referring to, "*the aggregation of this tracking across different websites into profiles and through attempts at linking this profile to the user's identity. By tracking these identifiers across websites that users visit, advertisers can infer users' interests, perhaps sensitive ones, such as medical conditions, political opinions, or even sexual fetishes".* To note, while users may be able to avoid some tracking by deleting cookies, persistent tracking technologies using unique identifiers now exist, also known as 'device fingerprinting'. They are made up of information recorded from a user's computer (e.g. from the specific browser version, plug-in, or other device configuration) which in combination may be enough to distinguish the device uniquely. Websites can use this fingerprint to attempt to recognise returning users. There are limited opportunities for users to prevent being tracked online in this way unless they use anonymity tools, like Tor ('The Onion Router') software, requiring a high level of technical proficiency to implement.

[13] Transparent Lives (Surveillance in Canada. Trend 4 – The Growing Ambiguity of Personal Information, [online]).

[14] Ibid: "*[t]he mobility of our devices means that we are continually connecting to the Internet at coffee shops, airports, and other public places through a number of IP addresses. Although an IP address is rarely going to be directly related to one identifiable individual, it is how the IP address is combined with other information (or could reasonably be combined with other information) about tastes, behaviours, and interests that has privacy advocates concerned. If you knew and combined enough online and offline information, you might have enough data to make a highly probable (sometimes almost perfect) guess about who was doing what, when, and where. For a simple scenario, consider the online consumer who leaves the IP address of his computer in access logs at each website visited. At some websites, the consumer may also provide explicitly identifying information; for example, his name and address are provided to complete a purchase. Separately, these websites can share logs containing the IP addresses of those who visited their sites. As e-businesses, these websites can also share explicitly identified data such as customer lists, which typically includes the name and address (e.g. residential, email, etc.) of those who made purchases. By examining the trails of which IP addresses appeared at which locations in the de-identified data and matching those visit patterns to which customers appeared in the identified customer lists, IP addresses can be related to names and addresses. These re-identifications can then be used to identify visits to locations in which the consumer did not make purchases".*

[15] While there are many definitions of big data, most reflect the growing technological ability to gather, aggregate, and process an ever-greater volume, velocity, and variety of data, including data collected in real-time. Other big data

provides another reason for significantly increased risk levels related to the prospect of being identified from information.

Analytics gain insight from data, and one of the most valued contexts of application for businesses and government is associated with the monitoring of human behaviour (collectively and individually). Constant, invisible tracking of online activity – an essential revenue stream for many companies doing business online - drives the growth of massive personal data sets. Looking for new insights from such data, analytics offer tantalising opportunities for organisations to re-use and thereby extract untapped data value, in turn increasing incentives to collect and retain as much data as possible, for as long as possible in case such opportunities materialise for future harvesting.

Advances in analytical processing technologies, improved data storing, machine learning, and the availability of large data sets including different data types, provide opportunities to link disparate aspects of data related to a single person in ways previously impossible. Data are gathered about individuals and analysed across large datasets ('mined') to identify correlations and facilitate people-'profiling' (i.e. automated forms of data processing for evaluating personal aspects - especially to analyse, classify, and make predictions about individuals - e.g., behavioural mapping).[16] Specifically, patterns of group behaviour ('class profiles' generated in respect of members with qualities appearing statistically-similar) can be applied to other individuals' profiles to infer new information about them (in particular, to identify characteristics and behaviour patterns), and ultimately potentially used to inform decisions taken in respect of them. All such processes may be carried out potentially without their knowledge.[17]

1.2.2 Identification capabilities are increasingly complex due to data linkage

These three trends point to a world in which it is now more easily possible to identify individuals through cross-correlation of datasets containing data indirectly about them. High-profile examples have been put forward to demonstrate that re-identification of individuals from data stripped of obvious identifiers may yet be possible through linking non-obvious features of such data with other

characteristics have been suggested including, veracity (if data may be inaccurate or unreliable), and variability (data with qualities that change, e.g., over time). According to Einav & Levin (2013, p.3): "*data is now available faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available*".

[16] UK Government, 2014, Emerging Technologies: Big Data - A Horizon Scanning Research Paper, p.2: "*[b]ig data refers to both large volumes of data with high level of complexity and the analytical methods applied to them which require more advanced techniques and technologies in order to derive meaningful information and insights in real time…Within this definition, there is a fundamental assumption about the power and importance of new techniques and technologies, which are often called 'analytics'.* ***The real value of analytics is that it can draw out new meanings, insights and value from bringing together individual datasets, which on their own might have limited value*****"** (emphasis added).

[17] On social media, for example, data analytics may be used on online-user data to create tailored real-time adverts from comparing individuals' specific browsing history and profiles with aggregated data on what consumers with similar profiles have purchased.

information containing identifiable data.[18] Facilitating such outcomes are organisations processing/sharing data relating to persons for secondary purposes not strictly necessary to achieving the original data collection purpose.[19]

Some argue that the tracking and processing of machine-generated information which is only obliquely about people – such as IP addresses - is not problematic since it identifies a device, not individuals. However, as mentioned, analytics can enable data linkage to other associated information facilitating personal identification. Indeed, the scale of data that can be collected across devices in aggregate causes the greatest problem due to the potential granularity of such detail and thus its sensitivity for its subjects, while additional information accretes with longitudinal tracking, increasing identificatory potential.

In this context, a distinction may be made between so-called 'identity disclosure', versus 'attribute disclosure', both of which are possible because of analytic profiling of individuals:[20]

- **Identity disclosure** refers to determination of a person's identity from information in combination with other information (i.e. based on a combination of facts associated with that person even if they do not include obvious identifiers). It is also possible to capture a host of data associated with a particular person, the analysis of which could enrich confidence in an identity determination.

- **Attribute disclosure** is another important outcome of analytics often associated with privacy risks. It means discovering something new and potentially sensitive about an individual from data, attributable to them without necessarily knowing who they are. This might involve the ability to infer something about an individual from information in a dataset(s).

[18] Ohm (2010), referring to studies involving re-identification of Google search histories, Facebook Friends lists, and Netflix film ratings. Another example is the AOL search results release incident, whereby two New York Times reporters were able to determine correctly the identity of one AOL user whose online search queries had been anonymised and posted on an AOL research group. For more information on anonymisation and re-identification, see Chapter 5.
[19] UK Information Commissioner (2015, pp.3-4): "*[a] key characteristic of big data is the ability to bring together data of different types and use it for new purposes. In some cases this may challenge the principle that personal data collected for one purpose should not be used for a further, incompatible purpose. Data protection requires that organisations process only the minimum data needed for a particular purpose. However, one of the driving forces behind big data is to collect as much data as possible, and may incentivise organisations to keep long runs of historical data. Organisations must still be able to justify this retention. …The propensity of big data to push against the boundaries of data protection is not an abstract issue. In certain applications it means increased risks of discrimination through profiling, of intrusion into private life, and of people being subject to decisions on the basis of processing that is opaque to them. Different risks arise from different big data scenarios - use for research purposes, to learn about trends and broader patterns, presents different data protection risks compared to when profiles are generated and used in a targeted way on individuals. Compliance solutions can be proportionate and matched to these different risks.*"
[20] See, e.g. El Emam (2013).

Notwithstanding this distinction of concepts, more attribute disclosure can ultimately lead to identity disclosure (if unknown) through the addition of informational links to a singled-out person.[21] Conversely, invariably accompanying identity disclosure is the discovery of more information (such as publicly-available online information) the analysis of which in combination may generate further attribute disclosure. Thus, possible protection-requiring interests resulting from information analysis from which personal identification might be possible is not limited to one type of disclosure risk, but involves a mix of both types.

1.2.3 Identification capabilities are a long-tail of possibilities and unique combinations

A discussion of non-legal (or not necessarily legal) meanings related to identification-associated terminology is set out in Appendix 1 below, as a complement to the legal analysis of such terms in Chapter 3. From this and the discussion above, it is evident that we can conceive the ability to identify someone from data – identifiability - as a spectrum of possibilities reflecting the different degrees of certainty capable of being associated with different identity decisions: the degree of likelihood of making an accurate identity decision. In practice, different degrees of confidence in the reliability of identity assurance decisions are socially acceptable. Said otherwise, there are different (margins of) error rates that are deemed acceptable depending upon the social context.[22] Furthermore, in this sense, identifiability can also often be linked to an objective conception of identity verification that is something which is either true or not as a matter of objective reality. For example, there is a tendency to treat biometric data as stable objects with well-defined and predictable outcomes.[23]

Contrariwise, identifiability also has a subjective quality in the sense that it depends on the one doing the identifying and the conditions of observation, i.e. the perception/cognitive processing of the information-expression of many different types of identifiers can vary.[24] Indeed, identification

[21] A theoretical question that flows from this distinction is how do we identify the point at which enough information elements can be accreted around a piece of data, such that it enables a particular individual associated with that data to be deemed identified (or identifiable, which implies a sufficient amount of probability that it could be done). For example, persons depicted on CCTV footage may not be immediately identifiable, but some may be made so fairly easily in combination with other associable information publicly-available (e.g. a social media profile photo). Chapter 3 discusses this issue further.

[22] For example, compare the level of proof to which you are put in getting a public library card with the degree of paperwork required to apply for a passport.

[23] For these reasons, the processing of biometric data is typically accompanied by specific safeguards enshrined in law (for further discussion, see Chapter 3).

[24] In particular, this is in light of the fact that a person will not necessarily reveal the same identifiers in the same way to different people. In this way, context and technical means can influence the ways in which an individual's identity may be revealed. Furthermore, whether a representation about one person serves to identify them depends on the audience's knowledge of and relation to this person. Returning to the biometrics example, Article 29 Working Party (2012, WP193, pp. 3-4) states: *"biological properties, behavioural aspects, physiological characteristics, living traits or repeatable actions where those features and/or actions are both unique to that individual and measurable,* ***even if the patterns used in practice to technically measure them involve a certain degree of probability"*** (emphasis added***).*** That

determination appears based upon a triadic relationship involving a person, an identifier (one or more, the meaning of which is discussed in Appendix 1 below) representing the person, and an observer processing (interpreting/ascribing meaning to) the relationship between the identifier in its representation and the person. However, as also described in Appendix 1, data combinations (specifically, uniquely-identifying combinations – 'strings of attributes' - produced by data mining analytics) enable 'quasi-identifiers' (viz. indirectly identifying variables) to arise. In this sense, the long-tail of identifiability risk extends way beyond the traditional domain of (obvious and non-obvious) identifiers of data subject.[25]

As an exercise in assessing possibilities, determining data identifiability is a difficult task considering that all information relating to us may give clues to who we are. Additionally, the more data points processed about us, the more person-inferences possibly generated algorithmically, and potentially the more accurate these may be. Further complicating this exercise are possibilities for combining data with other information together making identification potentially achievable where otherwise it may not be. Yet different people may assess the likelihood of identification from different perspectives (and positions of knowledge) in each case, so much uncertainty can persist in practice.[26] Future possibilities of an individual becoming identifiable resulting from new data combinations, or with emerging technologies, adds more complexity. Crucial additional information from unknown sources may become available sometime that could facilitate identifiability by someone somewhere.

At the same time having clear and coherent data protection rules is essential for continuous growth and confidence in the data economy, such as with the development of the Internet of Things (IoT). The rules should be strong but also flexible enough to encourage trust when organisations share data about people between them for socially beneficial outcomes, in an age when datafication, advanced data analytics and automated decision-making herald data-utility creation in new and innovative ways.

---

is, reliability varies intra-metrics according to the clarity of the images and the power of the technology available to perform this task. To this end, as pointed out by the House of Commons Science and Technology Committee (2015) regarding the scientific foundations of biometrics systems: *"[b]iometric recognition is a "probabilistic science"… biometric systems are affected by "intra-class variations". These are differences between two templates of the same trait, from the same user, captured at different times. Intra-class variations arise from multiple sources: "body parts age, sensors get grimy, lighting conditions change", all of which can introduce discrepancies between the same user's biometric templates. Biometric systems have to tolerate this degree of variability which, in practice, raises the prospect of false accept, and false reject, errors. In theory, the "science of biometrics" focuses on examining, and ultimately minimising, these errors to minimise the prospects of false accept, and false rejects, whether accidentally or malevolently committed".*

[25] Effectively, any attribute can be identifying (and uniquely identifying) in combination with others depending on the particular data context.

[26] Although if identification is construed as singling out as a minimum, this suggests that a more binary decision may be made: either they are, or they are not, capable of being singled out from the data.

In Part II, the research strategy is set out for successive chapters centred on how data protection law treats the identificatory requirement/element as part of the test for determining whether personal data exists (the processing of which, wholly or partly by automatic means, would result in the law applying).

Part II – The Thesis Research Approach

1.3 Research need

The preceding discussion raises two interlinking questions:

- To what degree should personal identifiability from data (as a state of the possible) be required to exist such that its processing merits legal data protection?
- How is expression of this degree conceptualised in law?[27]

Consideration of these questions are set against the proposition that – as argued in Chapter 2 - the identificatory requirement is meant to be the key delimiting legal factor in the personal data definition (scoping the outer boundaries of when data protection law applies to data processing).

When probing such issues, it is important to ensure that the law not only keeps pace with new technologies affecting interpretation of that requirement, but also that incorrect claims to satisfying that requirement (intentionally or unintentionally) are minimised as far as possible under the theoretical framework and in practice. Finding the identificatory requirement unsatisfactory in such respects would suggest an alternative requirement needs developing to underpin/delimit the legal concept of personal data.

1.4 Research objectives

A core objective of this thesis is to examine the identificatory requirement under data protection law, and to evaluate the proposition that it is no longer fit for purpose in light of modern technological advances. To this end, it seeks to do the following:

- Analyse the legal authority and theoretical framework associated with the identificatory requirement and its modelling (using an **'identificatory-approach'** to personal data).

[27] Conversely, how does the law conceive issues such as residual identifiability from data, the existence of which in practice would *not* be deemed sufficient to trigger the application of data protection law to such data (relating to persons) upon its processing?

- Consider the addressing of pinpointed shortcomings of the identificatory-approach using an alternative ('**effects-based**') approach to personal data, offering another way to define the outer boundaries of the concept under data protection law. Said otherwise, this approach – for development - would require assessing data processing *effects* as an alternate definitional requirement regardless of data's capacity to identify. Specifically, the *extent* of potential effects of processing of particular data on an individual would be the guiding legal principle, with personal data only encompassing information likely- capable of (at least) *appreciably* adversely affecting the individual to whom it relates.

- Consider whether, on balance, there are strong arguments for changing the law to an effects-based approach to personal data and consider models of the same.

Interrogating these different approaches will illustrate some of the practical advantages/disadvantages associated with different theoretical bases underpinning the personal data concept from a regulatory perspective. This could inform possible redefinitions of this concept looking beyond the 2018 changes to EU data protection law (see Chapter 2) by proposing a new theoretical approach to resolving the conceptual question - 'what are personal data?' - as a matter of legal/regulatory policy.

1.5 Research questions

The central thesis question is:

> **Would replacing the identificatory requirement for data to be deemed 'personal' under EU data protection law with a definitional provision linked to the assessment of effects flowing from the processing of data better comport with the twin data protection goals of facilitating the free flow of personal data, and safeguarding individual rights?**

Three sub-questions are proposed:

1. **Under the existing identificatory-approach to personal data (exemplified by the statement, 'information can only be personal data if it is capable of identifying the individual to whom it relates'), how effective is this approach in terms of realising and reconciling the twin goals of facilitating the free flow of personal data, and safeguarding individual rights?**

2. **Under an effects-based approach to personal data (exemplified by the statement, 'information can only be personal data if it is processed in a manner capable of affecting the individual to whom it relates appreciably'), how effective could this approach be in**

terms of realising and reconciling the twin goals of facilitating the free flow of personal data, and safeguarding individual rights?

3. **Can the practical disadvantages associated with an effects-based approach to personal data be ameliorated by the use of block exemption provisions, as exemplified under EU competition law (a distinct area of law and regulatory system that has been modernised using an effects-based policy approach)?**

To answer these questions, this thesis will:

- Analyse the legal concept of personal data and the identificatory requirement under existing EU data protection law - together with UK data protection legislation as an example of its implementation by an EU Member State (hereafter 'MS') - as well as under incoming EU data protection legislation. UK law has been chosen as exemplar in this respect because the UK is where the thesis author is based. Moreover, exploring UK data protection law in light of its EU origins helps makes better sense of the former.

- Develop a coherent, alternative approach that could underpin the legal definition of personal data.

- Explore the assumptions and conceptual legacies underlying both approaches and how they can give rise to different interpretations and models.

- Analyse each approach's implications for effective data protection regulation.

- Develop different models that could implement an effects-based approach of personal data to highlight and promote its potential benefits for effective data protection regulation at a practical level compared to the status quo (and the incoming status quo from 25 May 2018).

1.6 Outline of thesis contributions

Although this thesis advocates reconceptualising the definitional foundations of the personal data concept around a standard linked to assessing risks around likely effects from data processing, an existing body of literature on the topic of risk and its relationship with the definition of personal data already underpins much scholarly discussion, albeit one currently grounded upon an identificatory-approach.

Illustrative supporters of an identification/re-identification risk-based approach include, most notably, the following authors:

- **Ohm** suggests removing completely the personal data definition in law because of the (re-)identification risks involved when processing any data relating to persons; [28]

- **McCullagh** suggests that the question of when data is personally identifiable should be answered using a risk of (re-)identification approach;[29]

- **Tene** suggests restricting the scope of the term personal data based on the likelihood of identification with a context-specific test for assessing risk; [30] and,

- **Schwartz/Solove** suggest a risk of (re-)identification approach primarily considered under US law, although some general ideas/principles are also applicable to the EU legal context.[31]

In distinction from a repetition of these commentators' ideas, this thesis focuses on a definitional/jurisdictional framework that incorporates an assessment of the risk of effects – both negative (i.e. adverse impact, also known as harm) and positive, and potentially reflective of both individual and collective interests - that might flow from the implementation of a data processing activity. Implemented practically, the thesis' main ideas would incentives organisations intending to perform processing activities on information relating to persons to think about the risk of harm that might flow to such individuals and potential impact on other collective interests (e.g. affecting group privacy). Scholars who include consideration of likely harm into models of personal data are few and, moreover, their research approach can be distinguished from that set out in this thesis.[32] Whereas the preferred approach put forward in this thesis, in being able to take the collective interest into account, would be more pragmatic than existing alternatives, e.g. in emphasising an appropriate balance between the value of an anonymised dataset for secondary analysis and avoiding the likelihood of appreciable harm that might flow from the analysis felt at a group level (when the identities of members of the group are irrelevant).

The proffered approach would also be better suited practically to deal with challenges associated with new and complex technologies engineered with massive data exchanges between objects and big data analytics in mind (such as the IoT referred to above). In many cases the data being used for the analytics has been generated automatically, for example by tracking online activity, rather than being consciously provided by individuals. In that context, there is often much uncertainty about whether information qualifies as personal data including when information about people is

[28] Ohm (2010).
[29] McCullagh (2009, pp.13-24).
[30] Tene (2011).
[31] Schwartz & Solove (2011). See also Schwartz & Solove (2013).
[32] As discussed below, where scholars Hon et al (2011), and Gratton (2013) incorporate discussion of likely harm into their model of personal (or non-personal) data/information, they do so within the parameters primarily of an identificatory-approach to that concept, rather than under the primary 'banner' of a new approach.

transformed and re-used for a purpose different from that for which it was originally collected, and where the data may have been supplied by a different organisation. Yet such sharing and subsequent analyses can generate substantial social benefits, if exercised with sufficient caution, although assessment of such benefits are still at a relatively early stage (e.g. in relation to the development of machine-learning mechanisms for detecting patterns and making predications for earlier, automated detection and prevention of diseases).

In other words, the assessment of negative effects is only one side of the analysis. The other side is the assessment of positive effects and the inadequacies of the current legal framework calls for more effective tools to recognise this in a way that they can become embedded into the legal narrative. So it is crucial for organisations to be clear about the steps they have taken to address privacy risks and the potential benefits of what they are doing. Indeed, it will be argued that the capturing of such benefits could be expedited if there was the option of obtaining automatic legal comfort around there being an appropriate/proportionate level of protection guaranteed to such kinds of processing activities taking into account evidence of significant economic/social benefits capable of being achieved. This should also facilitate data sharing, while retaining clear accountability.

Since laws (their form, evolution upon revision, and guidance around their practical application) do not exist in a vacuum, it will be argued that development of an alternative approach could borrow from another regulatory regime that underwent effects-based modernisation in the early 2000s (EU competition law). Specifically, effects-assessment legal/regulatory tools adopted in this parallel area – notably block exemption ('safe harbour') regulations – might be moulded to an effects-based approach under the data protection regime to help resolve some practical challenges that otherwise remain outstanding. They will allow organisation to apply the proportionality principle as a guiding factor in deciding whether they should move forward with their data sharing proposals. They could also trigger considerations around ensuring ongoing and equitable benefit-sharing, as described below in Chapter 6.

1.7 Thesis structure

The thesis structure is as follows:

- **Chapter 2 –** contains an overview of EU and UK data protection law. It also describes its twin aims, together with the main sub-requirements composing the legal personal data definition.

- **Chapter 3** – examines legal interpretations of the identificatory requirement using related policy guidance and jurisprudence, and addresses sub-research question 1.

- **Chapter 4** – sets out the so-called Effects-based Approach for determining when data relating to persons may be classified personal data. It examines the legal heritage of this approach under existing data protection law and related EU and UK policy guidance, and addresses sub-research question 2.

- **Chapter 5** – splits into two parts. Part I concludes on the strengths and weaknesses associated with each approach, enabling a comparison to be made between them as to their effectiveness in light of the twin aims of data protection law. Part 2 springboards from such conclusions to the proposition that a new nuanced effects-based approach would be more realistic: one based on an exemption model. This is illustrated in the context of considering anonmyisation (techniques that are applied to personal data which may lead to them no longer being deemed personal) under data protection law and secondary sharing/re-usage of data concerns.

- **Chapter 6** – deals with sub-research question 3 in developing effects-based models based on exemptions, specifically to introduce block exemption regulations that require effects-centric analysis, permitting automatic legitimisation of certain aspects of data sharing/secondary analysis collaborations. Such regulations would not necessarily require the GDPR be amended to allow a carve-out, but they should provide a vehicle for stronger, more coherent effects-based theories in data protection law based on objective standards. The focus is twofold. First, the new models proposed in this chapter focus on reconciling the need for strengthened legal certainty under an effects-based approach model so far described (said otherwise, finding ways to mitigate levels of legal uncertainty raised as challenging in Chapters 4-5). Second, the model propositions focus on improving regulatory incentives for organisations processing data relating to persons to adhere to data protection rules pursuant to an effects-based approach, in turn encouraging high levels of organisational accountability and good practice in data management. The resulting recommendation is to introduce one or more proportionate effects-based block exemptions using secondary legislation, which would align in spirit and emphasis with incoming changes to the regulatory regime under the GDPR, but also enable the building of a new system of scalable data protection rules.

- **Chapter 7 –** concludes.

1.8 Research methodology

The research comprised a scrutiny of legislation as well as a library-based method consisting of a literature survey from secondary legal sources and books/journals. Aiding this review is interpretive guidance provided in case-law and by policy-makers.

Mostly, the research stems from considering material at a (pan-)EU level. However, to help understand the practicalities of implementing the EU definition of personal data, also considered is its interpretation and implementation in at least one EU MS (the UK), its national legislation, policy guidance, and case-law.[33] Furthermore, there are some references to concepts of personal data in the US and Australia (also known as 'PII' and 'personal information', respectively – see fn.2) under comparable data protection/privacy laws, where considered illuminative.

1.9 Research scoping

To enhance the accessibility of this discussion, some key research scoping decisions are worth mentioning briefly. These decisions have influenced the focus of this thesis, not least due to space constraints.

- First, it should be noted that – where reference is made to concepts of privacy and data protection - the focus is on how they are conceived in law, rather than exploring in any significant manner other disciplinary interpretations of such concepts (e.g. from sociological, or psychological, viewpoints). However, there is some critical analysis of literature in these other disciplines throughout the thesis (see in particular, Section 4.1.4 and Appendix 1) to reflect the Web Science nature of this thesis, whereby discussion is enriched through interdisciplinary insights.

- The legal perspective employed is primarily under data protection law, although briefly considered are legal rights to privacy and personal data protection, alongside reference to EU competition law as relevant to the analysis.

- Chapter 2 mentions the other requirements/elements for satisfaction under a legislative definition of personal data beyond the identificatory requirement. However, with the exception of the 'relating to' element, discussion of each is brief. Regarding the 'relating to' element - acknowledged to stand in close relationship to the identificatory requirement -

[33] A note on the impact of Brexit on this thesis is contained at Appendix 2.

for reasons explained in Chapter 2, the thesis does not engage in a detailed analysis of this element.

- Reasons given for comparing the two approaches chosen in this thesis do not preclude the existence of other possible ways of conceiving the personal data concept that might also be developed/compared.[34] However, this thesis will show that the main ideas put forward are notably better than such alternatives.

- Although the research underpinning this thesis commenced before the final text of the EU General Data Protection Regulation (GDPR, see Chapter 2) was agreed, it is considered alongside analysis of existing law. While this thesis could have focused solely on the new text as it mostly mirrors existing legislation (including in adopting an identificatory-approach to personal data and regarding the twin aims underlying both), it was deemed appropriate to consider both regimes in parallel. Notwithstanding, Chapter 6 highlights several new principles heavily-promoted under the GDPR intertwined with discussion of how to promote regulatory benefits post-2018 using an effects-based approach.

1.10 Introduction to case study for discussion in later chapters to illustrate this thesis' main ideas as they develop

1.10.1 Introduction

The purpose of this discussion is to describe a case study based broadly on the facts of a real ICO decision (the Decision) involving the processing and sharing of data relating to people between the Royal Free London NHS Foundation Trust ('the Trust', operating three London hospitals) and DeepMind (an affiliate of Google UK Limited). This case study is used to help make the thesis' original contribution clearer and highlight its uniqueness compared to other theories, in particular where an effects-based approach might diverge on application from an identification-based approach to data protection. To these ends, the case study facts will be developed hypothetically

[34] For example, in reminder, Booth et al (2004) conducted an analysis showing that EU MSs had different concepts of what constitutes personal data. They found that three conceptual approaches emerge: identificatory potential as a prerequisite; a relationship other than one of identificatory potential as prerequisite; and, identification and effect as prerequisites. However, their model of personal data based upon effect as prerequisite to data being personal was carried out in respect of interpreting the 'relating to' element of the legal definition of personal data under the DPD (see Chapter 2). It is therefore different from the current thesis study. Also noteworthy, Gratton (2013) proposes a hypothesis similar to this thesis author in suggesting an interpretation of whether data identifies or is identifiable of a person to fall within the 'personal information' concept (under Canadian law, as mentioned, broadly equivalent to the EU personal data concept) taking into account the underlying risk of harm associated with its processing. However, as also mentioned – see fn.32 above - she proposes *interpreting* this identificatory requirement *using* a risk of harm criteria, rather than *replacing* it altogether as this thesis author suggests. These are discussed in Chapter 4 below.

over Chapters 4, 5 and 6 to illustrate how changes to the modelling of an effects-based approach might result in different outcomes under existing law (and the GDPR), as well as any beneficial consequences of that difference, depending on the facts.

Such consideration of benefits may include the likely impact on the underlying information sharing collaboration between the Trust and DeepMind that might otherwise have been prevented following the Decision (although in reality, the project went ahead subject to substantial changes implemented following an audit per the ICO's requirements). Of course, there is also the possibility that the outcome is not better, or there is a mixed outcome of benefits and dis-benefits compared to the prevailing situation and the GDPR in counterfactual.

1.10.2 Background to the Decision

In 2015, Google UK Limited collaborated with the Trust to develop and deploy a new mobile app for the clinical detection, diagnosis and prevention of Acute Kidney Injury ('AKI') and an associated technology platform. The app – called 'Streams – hosts a NHS algorithm (a simple decision tree implemented across the NHS) that processes Trust-collected blood test results to identify patterns associated with AKI. If a pattern is found, Streams was designed to transmit a secure smartphone alert to an appropriate clinician's mobile, along with historical secondary care data about that patient (all information related to any treatment received by the individual from a Trust hospital in the last five years), to help them diagnose AKI and act immediately. The primary benefit associated with Streams was speed for a more efficient service (saving up to two hours of nurses' time per day), as well as providing opportunity to act quickly in life-threatening conditions.[35]

The Decision related primary to the data sharing aspect of the collaboration implemented following agreement of a data sharing contract between the parties.[36] The Trust sent over a million partial patient records to DeepMind encrypted in transit (via third party servers contracted by Google) in order that it could carry out clinical safety testing of Streams, and its underlying platform, before

[35] For more information on the facts, see e.g., Information Commissioner's Office. (2017). Royal Free - Google DeepMind trial failed to comply with data protection law. [online], 3 July 2017, available at: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/ [Accessed 1 August 2017]. See also King, D. & Suleyman, M. (2017). The Information Commissioner, the Royal Free, and what we've learned. [online] DeepMind, available at: https://deepmind.com/blog/ico-royal-free [Accessed 1 August 2017]. Notably, because Streams was designed to be ready for more advanced technology in the future, including AI-powered clinical alerts, the Trust hoped that it would help bring even more benefits to patients and clinicians in the long run.

[36] For more, see a critical examination of DeepMind's original data sharing agreement with the Royal Free Trust by researchers from Cornell University in Powles, J. and Hodson, H., 2017. Google DeepMind and healthcare in an age of algorithms. Health and technology, 7(4), p.351.

they became operational.[37] The Trust argued that it took a safety-first approach to testing Streams using real patient data to check the app was presenting the data accurately and working properly before going live.

The partial records shared included personal data (as acknowledged by the ICO), such as names, addresses, health service ID numbers, photographs and videos. Furthermore, the historical data would include the results of every blood test done at a Trust hospital prior to transfer, as well as all electronic patient records of admissions and discharges from critical care and accident and emergency. Consequently, that data could potentially include very sensitive personal information about individuals, such as the reason for admissions (e.g. a drugs overdose), or diagnosed conditions (e.g. HIV) that might or might not have a contributory significance to AKI.

The ICO's investigation was initiated to determine whether the Trust had complied with its data controller obligations under the DPA.[38] It concluded that the Trust was not fully compliant regarding the processing of personal data in this arrangement for a number of reasons:[39]

- Patients had not been adequately informed that their data would be used for the purposes of the clinical safety testing of the Streams app, and such usage would not have been reasonably expected by them.
- The personal data shared exceeded the minimum necessary for access by DeepMind relative to the purpose of enabling testing of the app.
- The Trust – who argued that it could rely upon implied consent of the patient as part of its primary care remit - had not demonstrated that it had a valid legal basis for processing the personal data, including sensitive health data, shared during the clinical safety testing phase.
- The processing was not subject to a full privacy impact assessment ahead of the project's commencement.

While the ICO did not ultimately impose a fine on the Trust (or DeepMind), it did require the Trust to sign an undertaking that it would carry out remedial action (e.g. improving transparency by way of providing additional information displayed on its website) following a third party audit.

---

[37] See, e.g., Denham, E. Letter to Sir David Solma, 3 July 2017, available at https://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf [Accessed 12 August 2017]

[38] Ibid - the ICO decided to assess DeepMind's practices on the basis that DeepMind was only a processor, acting on behalf of the Trust only.

[39] Ibid.

1.10.3 Why this case study?

The Decision is an interesting case on the facts. It involved the processing of sensitive health care data but in a context where no objective, appreciable negative effects to individuals were obvious as a result of the processing stage (i.e. the testing of the clinical app) taking into account existing measures implemented. For example, DeepMind's security measures were considered more than adequate and indeed almost certainly superior to the NHS's security arrangements. Conversely, the benefits of the processing stage were undeniable - to achieve the establishment of an innovative technological solution, Streams, and improve clinical care - for which the appropriate use of personal data was required. Thus, the facts highlight how both the risks of negative effects but also the potential delivery of positive effects in the public interest, might be worthy of consideration (implicitly or explicitly) flowing from a data processing arrangement if permitted under data protection law.

This harm/benefit dichotomy narrative is one that pervades recent discussions of big data analytics and machine-learning involving the processing of data relating to people. Benefits to contemplate include those to individuals, groups and at a societal level. For example, they can include helping develop better health care services that improve the quality of people's lives, and also improve patient treatment systems generally. Recognition of this raises the following question: how could the law better balance the socially beneficial uses of big data with the harms to privacy and other values that can result from particular processing uses?

At the same time, the question of whether personal data was processed (a jurisdictional issue) was not pivotal to the Decision: as mentioned, it was acknowledged that it was. However, in Chapters 5 and 6 below the case study facts are developed (hypothetically) to focus on the issue of anonymisation and consider its impact. As discussed here, the re-use of data relating to people, and the transformation of such data (through anonymisation techniques) to mitigate the risks arising from its secondary use, is an issue that continues to need addressing as big data analytics becomes increasingly a part of our daily lives in a range of economic sectors. A particularly important risk factor in many data sharing scenarios is the risk of re-identification of anonymised data, including by other organisations. This is because the public interest in re-using data extends to maintaining public trust when innovations with databases are taken forward, especially when they concern access to sensitive data (such as health care data), and could result in severe consequences around loss of trust.[40] Moreover, as mentioned, such concerns – and opportunities – are likely to increase

[40] An example of where public trust was lost relates to the aftermath of the launch of NHS England's care.data programme regarding electronic patient records in 2014. The programme was intended to collect primary care data from GP practices, and link this data with secondary care data (from hospitals, registries and prescribing databases)

in the future as linkage and reuse of electronic patient data, or whole genome sequencing, becomes part of standard clinical care driving advanced health informatics and clinical research.

Another aspect of this debate is how to provide data holding organisations with confidence and indeed incentives to disclose information to third parties for valuable data re-usages in compliance with data protection law (i.e. not held back by lack of legal certainty, including a lack of clarity about de-identification standards in law). Indeed, the ICO has published a blog – and the Information Commissioner – has directly spoken about the lessons that can be learnt from the Decision in such terms: "*As organisations increasingly look to unlock the huge potential that creative uses of data can have for patient care, what are the lessons to be learned from this case. It's not a choice between privacy or innovation …. The price of innovation didn't need to be the erosion of legally ensured fundamental privacy rights".*[41]

In that context, an effects-based approach may have an important role to play and constitutes a worthy avenue of investigation around how to increase trust and positively impact on data utility simultaneously. It could help promote secondary usage of data relating to people as part of the legal framework, and/or as part of regulatory toolkit encouraging compliance, in ways that encourage positive effects such as potential flowing from generalised analytic processes and demonstrate what compliance in particularly innovative and demonstrably publicly-beneficial ways look like. Ideally, it would also provide assurances to the public that there are strong safeguards in place to make sure that data relating to them will only ever be used transparently, safely and in line with the law and regulatory framework (i.e. to build confidence that data is being used for public benefit and in a responsible way, underpinned by trustworthy data governance processes).

Although an effects-based approach put forward in this thesis admittedly will not provide a complete solution to the issues arising in this case study and around anonymisation in general, it is hoped that the models proposed in later chapters could act as a departure point to develop the legal framework and play a role in widening benefits recognition, post-GDPR's introduction, in the future where data re-usage is increasingly characterised by digital innovations such as Streams.

---

nationally in order to advance research, provide better care, and inform commissioning. Following public backlash about the weak governance processes regarding data usage underpinning the programme and its inadequate communication conveying its benefits and safeguards, care.data was eventually abandoned. See also Royal Statistical Society, 2014. Royal Statistical Society research on trust in data and attitudes toward data use / data sharing, available at: https://www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf [Accessed 1 August 2017]. This report highlighted that there is a data trust deficit giving rise to missed opportunities for valuable and life-saving research.

[41] Information Commissioner's Office. (2017). Royal Free - Google DeepMind trial failed to comply with data protection law. [online], 3 July 2017, available at: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/ [Accessed 1 August 2017]. Elizabeth Denham, Information Commissioner, went on to say, "*The Data Protection Act is not a barrier to innovation, but it does need to be considered wherever people's data is being used*."

1.11 Chapter conclusion

This chapter highlights how with new analytical techniques, and more data recorded/stored, comes increased possibilities for identification from the information that gathers around us. Moreover, identification likelihood and privacy-invasiveness[42] has an accretive data nature: the more information revelation about a person, the easier it becomes to trace them in connection with that data, and make attribute/identity disclosures about them. Indeed, it is possible to argue that any attribute associated with an individual is identifying of them as long as the corresponding data attributes are sufficiently numerous to enable them to be distinguished uniquely from other persons in a dataset. In parallel, identifying someone from non-obvious online identifiers can happen now in myriad - increasingly technical – ways, by combining such data with other associated information, including via automated profiling activities often remaining hidden to the individuals involved.

While the exact meanings of identificatory concepts in a non-law setting are rather vague, a continuum of identification possibilities may be conceived to exist leading to the notion of personal identifiability from data as a spectrum, with various degrees between its extremes (of clearly identifiable from data relating to persons, to not identifiable at all). Understanding identifiability as a potentiality also entails consideration of roles – the object, and subject, of the verb 'to identify', as well as the data-embedded identifier(s) – but also motives and incentives for attempting identification, as these can all affect the likelihood of success. Yet, data protection law imposes obligations that require making binary identification determinations, whereby the regime is either applicable or not based on the results of such determinations. We turn to the main principles of EU and UK protection law and their legislative definitions of personal data, alongside discussion of the twin aims of data protection assessed against this backdrop.

[42] For a discussion of the concept of privacy, see Chapter 4. For the moment, it is referred to in the context of unwanted attribute disclosure about a person or persons in a particular context, not precluding alternate definitions in non-legalistic (or, indeed, myriad possible definitions of a right to privacy protection in legalistic) contexts.

Chapter 2 - Overview of EU and UK Data Protection Law

This chapter describes the legal framework underpinning the analysis in successive chapters, specifically EU data protection law and its implementation/regulation in EU Member States (hereafter 'MSs'), with the UK regime as exemplar. In terms of chapter structure:

- **Section 2.1** provides an overview of EU law related to privacy and data protection.
- **Section 2.2** summarises the relevance of this overview for analysing the research question/sub-questions (in terms of outlining the assessment framework for determining which model of personal data for review comports better with data protection's twin goals).
- **Section 2.3** describes the different elements of the legal definition of personal data, including introducing its identificatory requirement.
- **Section 2.4** concludes.

2.1 EU law protecting privacy and data protection

Primary and secondary laws provide EU legal standards for privacy and personal data protection. This section considers both briefly, followed by a closer examination of the background to the Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data (DPD) and its twin aims. This analysis helps understand the context of the research question and the underlying framework of assessment in readiness for applying from Chapter 3.

2.1.1 Primary EU law (fundamental rights as general principles of EU law)

The EU is an international organisation governed by treaties ratified by its MSs. These treaties form primary EU law. While there was a lack of specific and exhaustive provisions for the protection of fundamental rights in the founding EC Treaty,[43] with its primary goal of attaining economic integration by establishing a European Community Common Market, this has not meant an absence of legal protection.

[43] The Treaty of Rome, originally the Treaty establishing the European Economic Community (TEEC), succeeded by the Treaty establishing the European Community.

As early as the 1960s, the European Court of Justice (now Court of Justice of the EU, CJEU) held that fundamental rights and protections (hereafter, 'rights') form an integral part of the general principles of law whose observance it ensures.[44] The CJEU continues to interpret and review EU measures in the light of such rights to this day. They constrain not just the EU Courts (the CJEU and the General Court (formerly Court of First Instance)) but also MSs' courts. The latter are obliged to interpret/adhere to domestic laws in EU matters according to these rights and, accordingly, must disapply conflicting domestic law provisions. MSs must also not introduce laws deviating from them.

Modern-day EU treaties now enshrine explicitly a system of rights protection.[45] A treaty provision specifically dealing with this issue was introduced under the 1992 Maastricht Treaty (TEU) when a new Article F(2), essentially codifying the CJEU's case-law, provided that the EU must respect rights as general principles of law. Subsequent amendment to this treaty provision, latterly known as Article 6(2) TEU following the entry into force of the 1999 Treaty of Amsterdam, made clear that the EU is based, inter alia, on the principle of respect for rights (Article 6(1) TEU).

With the 2009 entry into force of the Treaty of Lisbon (also known as the Treaty on the Functioning of the EU, TFEU), it stated (at Article 6(3)) that, *"[f]undamental rights, as guaranteed … shall constitute general principles of the Union's law*". The TFEU also introduced a new legal basis for EU-wide rules applicable to all public/private sector processing of personal data (Article 16). In the same year, the adoption of a codified declaration of EU rights further strengthened the protection provided to EU citizens' interests: the Charter of Fundamental Rights of the EU (Charter).[46] The Charter consolidates various sources of EU rights' protection into a single document, and reaffirms their importance under EU law, e.g. with respect to the powers/tasks of the EU institutions, and vis-à-vis the 'principle of subsidiarity' regarding when it is preferable for action to be taken at EU-level, rather than by MSs. These sources include the European Convention on Human Rights (ECHR), national constitutional traditions, and international obligations of MSs, and the Council of Europe,

[44] See, Case 29/69, Erich Stauder v. City of Ulm – Sozialamt [1969] ECR 419 303.

[45] This fact is despite the EU not directly being a signatory to the European Convention on Human Rights (ECHR), a treaty adopted by the Council of Europe (an international organisation distinct from the EU and its institutions) to which each of the EU MSs belongs. Article 8 ECHR protects the right to respect for the private and family life of individuals. The European Court of Human Rights (ECtHR) that rules on matters covered by the ECHR has interpreted this Article to encompass the protection of the processing of personal data. The non-signatory status of the EU has not precluded the CJEU from referring to the case-law of the ECtHR when developing its fundamental rights jurisprudence via the notion of general principles of law (see, e.g. Joined Cases 46/87 and 227/88, Hoechst AG v. Commission [1989] ECR 2859). The CJEU also refers to the constitutional traditions of MSs, and international human rights treaties, to assist with its interpretation of EU rights' standards. For example, rights to privacy afforded to individuals are protected in international human rights instruments, such as Article 12 of the Universal Declaration of Human Rights, or Article 17 of the International Covenant on Civil and Political Rights.

[46] This was when the Charter became primary law. Article 6(1) TEU provides that the EU, "*recognises the rights, freedoms and principles set out in the Charter of Fundamental Rights of the European Union of 7 December 2000, as adapted at Strasbourg, on 12 December 2007, which shall have the same legal value as the Treaties*".

as well as the distinct case-law bodies of the CJEU and the European Court of Human Right (ECtHR). The Charter also codifies the right to personal data protection with the status of a right under EU law in its Article 8, as distinct from the right to respect for private and family life in its Article 7.

2.1.2 Secondary law (the DPD)

Secondary EU legislation are instruments such as regulations, directives, and decisions (with the differences between the first two types considered below).[47] The DPD is the main EU secondary law in the area of data protection law. As a directive, the DPD requires MSs to achieve particular results within their national laws through the implementation of its provisions, without dictating how precisely to achieve those results.[48]

2.1.2.1 Nature and scope of the DPD

The DPD introduced an extensive data protection regime applying in situations where specified trigger (so-called 'jurisdictional', or 'gateway-level' scoping) factors are met. As well as the applicability of data protection rules under the DPD being conditional upon data being personal, the data at issue must also be processed.[49]

When triggered, the DPD stipulates basic data protection principles that must be followed. Primarily, Article 6(1) requires MSs to provide that personal data must be:

> (a) processed fairly and lawfully;
>
> (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes…;
>
> (c) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;

---

[47] Alongside these legally-binding measures, non-binding measures with high legal influence may also be issued by EU institutions and associated bodies. For example, the European Commission can issue recommendation and opinions (such as in the form of guidance, communications, and notices). These can be used to provide context to and interpret binding instruments. The latter function can be particularly useful where binding law lacks clarity, such as where its provisions are drafted in ambiguous language. Examples are discussed further below.

[48] In other words, it is merely binding as to the result to be achieved upon each MS to which it is addressed but shall leave up to national authorities the choice of form and methods for implementation through domestic legislation to be adopted.

[49] Article 2(b) DPD defines 'processing' broadly as, "*any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organisation, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction*". Article 3(1) states that the DPD applies to, *"the processing of personal data wholly or partly by automatic means*" and *"the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system*". According to Article 2(c) DPD, a 'personal data filing system' is *"any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical bas*is".

> (d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified;
>
> (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed...;

Data controllers must ensure that these principles are complied with (Article 6(2)), and take steps to ensure the confidentiality/security of personal data processing (Articles 16-17). Besides imposing broad obligations on 'data controllers',[50] the DPD confers broad rights on individuals about whom data are processed ('data subjects'). The entities that process their personal data must inform them how it will be used, and to whom it will be transferred. Data subjects also have rights of access to their personal data, encompassing rights to:

- know if data is being processed about them;
- a description of certain matters (e.g. the recipients to whom the data may be disclosed);
- have certain information communicated in an intelligible form;
- object to data processing, and rectify their personal data, in specific circumstances;
- be informed of the logic of automated decision-taking being taken about them; and,
- not be subject to certain decisions made solely based on automated data processing.

The DPD specifies the grounds on which the processing of personal data can be justified (also known as 'legal bases'), such as where the data subject's consent has been obtained.[51] Moreover, data

[50] Articles 2 (d) and (e) DPD define the concepts of data controller and data processor, respectively. The definition of a data controller is, "*the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific criteria for his nomination may be designated by national or Community law*". A data processor, by contrast, is defined as "*the natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller*". For ease, in this thesis, the terms 'data controller' (or, occasionally, 'data holder') are used broadly to denote an entity that determines the purposes and means of the processing of data relating to persons broadly (which may or may not be personal data upon further assessment). Similarly, the term 'data subject' is used broadly to denote someone about whom data for processing is related (which may or may not be personal data upon further assessment).

[51] Article 7 DPD. Personal data should only be processed if: *"(a) the data subject has unambiguously given his consent; or (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or (c) processing is necessary for compliance with a legal obligation to which the controller is subject; or (d) processing is necessary in order to protect the vital interests of the data subject; or (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed; or (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or*

transfers to non-EU countries are only permitted if the recipient country's applicable laws guarantee an "*adequate level of data protection*" (Article 25).

2.1.2.2 Origins and background to the DPD and its twin aims

The DPD's basic concepts were modelled on the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108), an international treaty for data protection adopted by the Council of Europe's MSs in 1981. Convention 108 acknowledges the need to, "*reconcile the fundamental values of the respect for privacy and the free flow of information between peoples*". Such recognition is motivated by a desire to achieve greater unity between such MSs and "*extend the safeguards for everyone's rights and fundamental freedoms, and in particular the right to the respect of privacy, taking account of the increasing flow across frontiers of personal data undergoing automatic processing*".[52] Convention 108 can be seen as a prelude to the DPD and its concern for the protection of the rights of EU citizens, notably their right to privacy.

The DPD was adopted under Article 95 EC (of the EC Treaty), the legal basis for adopting legal measures related to the achievement of a Single Market within the European Community (the EU's predecessor).[53] Specifically, the intention behind the DPD was to facilitate the harmonisation of MSs' domestic laws on data protection across Europe and prevent impediments to the achievement/maintenance of a unified trade bloc (Common or Internal Market, later evolved into the more ambitious concept of the EU Single Market).[54]

To understand more fully this rationale, we turn to the DPD's first Article. Article 1(1) lays down its object by stating that, "*in accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data*". Article 1(2) states, "*Member States shall neither restrict*

---

parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1)".

[52] Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, (1981), at preamble: "*[c]onsidering that the aim of the Council of Europe is to achieve greater unity between its members, based in particular on respect for the rule of law, as well as human rights and fundamental freedoms; Considering that it is desirable to extend the safeguards for everyone's rights and fundamental freedoms, and in particular the right to the respect for privacy, taking account of the increasing flow across frontiers of personal data undergoing automatic processing; Reaffirming at the same time their commitment to freedom of information regardless of frontiers; Recognising that it is necessary to reconcile the fundamental values of the respect for privacy and the free flow of information between peoples".*

[53] Article 95 of the EC Treaty provides that the Council shall adopt measures for approximation of provisions laid down by law, regulation or administrative action in MSs which have the objective of establishment and functioning of the Internal Market. It may be used where disparities exist or potentially exist between MSs' laws that obstruct fundamental freedoms or create distortions of competition. In other words, the spirit of the principles that applied to the free movement of goods, services, persons, and capital under Article 7a of the EC Treaty (and the prevention of obstacles being applied thereto by MSs) was conceived to apply also to personal data.

[54] In addition, the DPD applies to the three European Free Trade Area (EFTA) Member States who are also members of the European Economic Area (EEA) but not members of the EU (i.e. Iceland, Norway, and Lichtenstein). References in this thesis to 'Member States' and 'EU' are intended to also include those countries.

nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1". Whereas one explicit aim of the DPD, therefore, is to support the necessary conditions for the establishment/functioning of the Single Market - ensuring the free flow of personal data across MSs (see also Recital 5)[55] - it also had another objective. This second aim is to guarantee the protection of individuals' rights – including, but not limited to, the right to privacy - with respect to the processing of their personal data (see also Recital 10).[56]

The DPD's Preamble further explains how these twin aims relate. It alludes to European decision-making bodies adjudging that differing levels of rights' protection for individuals afforded in different MSs constitutes an "*obstacle to the pursuit of a number of economic activities at Community level*" (Recital 7), [57] inhibiting personal data movement cross-border. The approximation of national laws could overcome this obstacle (Recital 9).[58] Therefore – in aiming to ensure consistent regulation of cross-border personal data flows aligned with Single Market objectives - the DPD obliges MSs to realise an *equivalent (and high)* level of protection of individuals' rights when their personal data are processed (Recital 8).[59]

---

[55] Recital 5, DPD: "*[w]hereas the economic and social integration resulting from the establishment and functioning of the internal market within the meaning of Article 7a of the Treaty will necessarily lead to a substantial increase in cross-border flows of personal data between all those involved in a private or public capacity in economic and social activity in the Member States; whereas the exchange of personal data between undertakings in different Member States is set to increase; whereas the national authorities in the various Member States are being called upon by virtue of Community law to collaborate and exchange personal data so as to be able to perform their duties or carry out tasks on behalf of an authority in another Member State within the context of the area without internal frontiers as constituted by the internal market*".

[56] Recital 10, DPD: "*[w]hereas the object of the national laws on the processing of personal data is to protect fundamental rights and freedoms, notably the right to privacy, which is recognized both in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms and in the general principles of Community law; whereas, for that reason, the approximation of those laws must not result in any lessening of the protection they afford but must, on the contrary, seek to ensure a high level of protection in the Community*". In that context, to note, while the concepts of data protection and privacy are closely related under the DPD, they are not synonymous (a fact discussed further in later chapters).

[57] Recital 7, DPD: "*[w]hereas the difference in levels of protection of the rights and freedoms of individuals, notably the right to privacy, with regard to the processing of personal data afforded in the Member States may prevent the transmission of such data from the territory of one Member State to that of another Member State; whereas this difference may therefore constitute an obstacle to the pursuit of a number of economic activities at Community level, distort competition and impede authorities in the discharge of their responsibilities under Community law; whereas this difference in levels of protection is due to the existence of a wide variety of national laws, regulations and administrative provisions".*

[58] Recital 9, DPD: **"***[w]hereas, given the equivalent protection resulting from the approximation of national laws, the Member States will no longer be able to inhibit the free movement between them of personal data on grounds relating to protection of the rights and freedoms of individuals, and in particular the right to privacy; whereas Member States will be left a margin for manoeuvre, which may, in the context of implementation of the Directive, also be exercised by the business and social partners; whereas Member States will therefore be able to specify in their national law the general conditions governing the lawfulness of data processing; whereas in doing so the Member States shall strive to improve the protection currently provided by their legislation; whereas, within the limits of this margin for manoeuvre and in accordance with Community law, disparities could arise in the implementation of the Directive, and this could have an effect on the movement of data within a Member State as well as within the Community".*

[59] Recital 8, DPD: "*[w]hereas, in order to remove the obstacles to flows of personal data, the level of protection of the rights and freedoms of individuals with regard to the processing of such data must be equivalent in all Member States; whereas this objective is vital to the internal market but cannot be achieved by the Member States alone, especially in view of the scale of the divergences which currently exist between the relevant laws in the Member States and the need to coordinate the laws of the Member States so as to ensure that the cross-*

In assessing the nature of the relationship of such twin aims, they may be seen as standing in contra-distinction. That is, more protection of individual rights under data protection law may be considered liable to have the effect of restricting free-flowing personal data inter and intra EU MSs (because more data relating to persons would become subject to data protection rules, in turn inhibiting data flows),[60] and vice versa.[61] However, an alternate interpretation may also be considered, whereby the adequate safeguarding of fundamental rights is a pre-condition to achieving the free flow of personal data. Specifically, it is only by coordinating data protection laws pan-EU to a high level of rights protection that the free flow of personal data can be achieved.[62]

This second, alternate interpretation is supported by evidence that the DPD was conceived as a legislative measure to foster each aim in a reconciliatory (rather than opposing) manner. For example, Recitals 2-3 refer to the creation of a high level of data protection throughout the EC as playing a key role in facilitating a free flow of information in support of the Single Market because of the consistency it engenders.[63] Other support for giving both twin aims intertwining and equal importance can be found in the CJEU's *Lindquist* judgement.[64] The anticipation by the CJEU of the harmonisation of national data protection laws following the DPD's implementation in MSs' laws was envisaged as ensuring a high level of rights protection, in turn facilitating personal data flows throughout the EU.[65] As such, it appears that the legislative effort to bring about a practical concordance between the DPD's twin aims was envisioned whereby neither of them takes precedence unilaterally over the other.

---

border flow of personal data is regulated in a consistent manner that is in keeping with the objective of the internal market as provided for in Article 7a of the Treaty; whereas Community action to approximate those laws is therefore needed".

[60] For example, because consent or another recognised legal basis (i.e. processing ground justifying personal data processing taking place) would have to be in place before such data could be processed, in turn involving effort and potentially legal spend to ensure compliance, the effect of which could be to hinder the free flow of such data.

[61] For example, because – the less the reach of data protection law (in reflecting a lower standard of privacy protection) – the less effort and expense that must be expended in ensuring data protection law compliance by organisations, which could encourage them to engage in more personal data exchanges and increase the flow of such data.

[62] This has the implication that too much rights protection may mean a restriction of the free flow of personal data across MSs *only where* it is at the expense of consistency/equivalence across the law in these territories.

[63] Recital 2, DPD: "*[w]hereas data-processing systems are designed to serve man; whereas they must, whatever the nationality or residence of natural persons, respect their fundamental rights and freedoms, notably the right to privacy, **and** contribute to economic and social progress, trade expansion and the well-being of individuals*" (emphasis added); Recital 3 – "*[w]hereas the establishment and functioning of an internal market in which, in accordance with Article 7a of the Treaty, the free movement of goods, persons, services and capital is ensured require **not only that** personal data should be able to flow freely from one Member State to another, **but also that** the fundamental rights of individuals should be safeguarded*" (emphasis added). This interpretation of the relationship between these twin aims as complementary can also be gleaned from Article 16 of the TFEU (ex Article 286 TEC), which immediately after the edict that "*[e]veryone has the right to the protection of personal data concerning them*", provides that the EU's legislature shall "*lay down the rules relating to the protection of individuals with regard to the processing of personal data…and the rules relating to the free movement of such data*".

[64] Case C101-01, Bodil Lindqvist [2003] ECR I-12971.

[65] See also Joined Cases C-468/10 and C-469/10, Asociacion Nacional de Establecimientos Financieros de Credito (ASNEF) and Federacion de Comercio Electronico y Marketing Directo (FECEMD) v. Administracion del Estado [2011] ECR I-12181.

2.1.3 National implementation of the DPD

The DPD required MSs to adopt national law (by 25 October 1998) adhering to its objectives/principles in order to harmonise national data protection laws between them (Article 32(1)-(2)). However, as described above, as a directive, MSs had margin for manoeuvre in devising such laws in implementation of the DPD's broad legislative framework. Thus, each MS enacted their own domestic data protection laws, giving rise to national differences.[66]

One reason for such variations in practice is the fact that the DPD provides that MSs may exempt their laws from compelling certain obligations. Another reason, more relevant to this analysis, is the fact that many of the DPD's provisions are at points relatively general and ambiguous (in some respects unclear) since they apply widely, and it was left to MSs to construe (interpret/clarify) their meaning. These can make for significant differences in the operation of national data protection regimes. To this end, national data protection authorities (DPAs) – including the UK's Information Commissioner's Office (ICO), an independent regulatory body responsible for upholding information rights, including matters regarding privacy and data protection - regularly publish opinions and other documents dealing with data protection topics expanding upon their interpretation of the DPD's key provisions within their jurisdiction. Chapter 3 explores this issue further in respect of the DPD's identificatory requirement under the personal data concept as implemented in UK law.

2.1.4 National enforcement and regulatory oversight of data protection law

Under the DPD, supervision of fair and lawful personal data processing in the MSs must be ensured, including through establishing DPAs (at least one in each country) responsible for monitoring the application of domestic data protection laws within their jurisdiction.[67] Furthermore, each MS had to adopt suitable enforcement-related measures to promote compliance with national rules, including laying down sanctions that can be imposed by DPAs against non-complying organisations/individuals. These commonly include powers of investigation and intervention, as well as fining systems that may be applied to enforce infringement decisions once taken.[68]

[66] It also gave rise to potential allegations of mis-implementations of the DPD's provisions in national laws, such as have arisen in respect of the definition of personal data under the UK Data Protection Act 1998 as discussed later in this chapter.

[67] Under Article 28 DPD, alongside Article 8(3) of the Charter and Article 16(2) TFEU, supervisory authorities, acting independently, are to monitor compliance with EU provisions dealing with the protection of individuals with regard to the processing of personal data.

[68] Aside from legal sanctions, non-compliance can also result in damaging adverse publicity.

As the approaches taken by the DPAs in developing their data protection regulatory regimes could vary widely in practice between countries, a pan-EU working party was established under Article 29 of the DPD (the Article 29 Working Party, hereafter WP). This body is composed of representatives of the MSs' DPAs, the European Data Protection Supervisor (who regulates EU institutions' data protection compliance), and the European Commission. The WP regularly issues statements (opinions and other non-binding but highly-influential documents) on a range of data protection related topics, providing a useful body of guiding resource in interpreting DPD and domestic law (DPD-implementing) provisions.

2.1.5 EU data protection reform (the GDPR)

Despite two decades passing since the DPD was adopted, variances in interpretations of key data protection provisions between MSs is still seen as problematic, insofar as non-harmonisation of data protection laws across the EU hampers the achievement of the Single Market.[69] Inconsistent interpretations of the DPD's major provisions as implemented in national data protection laws arise between the DPAs, but also between DPAs and national courts. Accordingly, the EU has prioritised efforts to remove any residual legislative barriers to improve data protection legal harmonisation between MSs.

Partly for this reason, the Commission initiated a reform programme in 2012 when it proposed a comprehensive review into existing data protection rules in the EU, and published a draft General Data Protection Regulation (GDPR) for consultation to replace the DPD.[70] Other rationale included a desire to update/strengthen data protection rules in ways better suited to achieving a digital Single Market. In the Commission's draft GDPR text (COM/2012/011 final), for example, it proposed modernising and clarifying the application of some key DPD rules considering the impact of the advent of new technologies on the interests that underpin data protection principles. Recommendations contained therein also focused on reinforcing the rights of individuals,[71] as well as generally improving the effectiveness and coherency of the EU-wide regulatory regime.

[69] See, e.g. the research study by Booth et al (2004) commissioned by the ICO, which concluded – at least with respect to the period up to the early 2000s - that there was no coherent definition of personal data applied consistently pan-EU.
[70] Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) issued by the European Commission on 25 January 2012. To note, as part of the reform discussions, two other principal draft versions were issued by the European Parliament (a 2014 legislative resolution) and the European Council (a 2015 general approach). The GDPR was the subject of extensive debate and lobbying as well.
[71] Ibid, Recital 9: *"[e]ffective protection of personal data throughout the Union requires strengthening and detailing the rights of data subjects and the obligations of those who process and determine the processing of personal data, but also equivalent powers for monitoring and ensuring compliance with the rules for the protection of personal data and equivalent sanctions for offenders in the Member States".*

The legislative process to enact the reform concluded in 2016, with the definitive version of the GDPR adopted formally into EU law.[72] Its enactment repeals the DPD and thus supersedes national data protection laws that implemented the DPD throughout the EU. It comes into force following a two-year transition period on 25 May 2018. The deliberate change from a directive to a regulation, as the preferred form of harmonising legal instrument of data protection rules at EU level, marks a significant change.[73]

An EU directive is binding as to the results to be achieved upon each MS to which it is addressed requiring them to pass legislation to implement its standards into national law. As such, it potentially provides MSs with specific discretion in how they implement certain issues as they think fit and gives rise to inconsistent implementation between different MSs.[74] In comparison, a regulation is binding immediately without the need for transposition into national law within each territory.[75] Thus, a regulation can create one single, directly applicable law enforceable immediately in its entirety throughout the EU.[76] The introduction of a single legal framework under the GDPR, with uniform wording on data protection regulation EU-wide, should improve legal certainty.[77] It should also improve, in turn, the protection of individuals' rights and contribute to the functioning of the Single Market, which are both explicit goals of the GDPR (as with the DPD). Recital 7 GDPR states:

> The objectives and principles of Directive 95/46/EC remain sound, but it has not prevented fragmentation in the way data protection is implemented across the Union, legal uncertainty… Differences in the level of protection of the rights and freedoms of individuals, notably to the right to the protection of personal data, with regard to the processing of personal data afforded in the Member States may prevent the free flow of

[72] Regulation (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 Directive 95/46/EC (General Data Protection Regulation).

[73] In particular, data protection is no longer perceived as a local phenomenon, to be regulated according to local legislation with an EU directive only issuing high-level instructions and guidelines. Now data protection is considered as a supra-EU concern, to be regulated directly at EU level in a common manner for all MSs through a regulation (in the same way that, say, EU competition law is, as discussed further in Chapter 6). Notwithstanding, the GDPR contains a certain amount of so-called 'opening clauses' granting MSs some discretion as to how to they enact their domestic laws to specify the GDPR. Further, in some cases, MSs are obliged to provide for specifications on a national level.

[74] To note, moreover, a directive (or clauses within a directive) may be described as 'maximum harmonisation' meaning that domestic laws in implementation may not exceed the requirements of the directive, contrasted with 'minimum harmonisation' where national laws may impose additional requirements if they so choose. The DPD is typically considered to be a minimum harmonisation directive.

[75] According to Article 288 TFEU, an adopted regulation is binding immediately from the date specified in it (or, if unspecified, on the twentieth day following its publication in the Official Journal (OJ) of the EU) and incorporated into the domestic legal orders of all MSs.

[76] Notwithstanding, MSs may introduce domestic legislation to support the introduction of the GDPR as currently being considered in the UK with the introduction of the Data Protection Bill 2017.

[77] As Kuner (2012, p.3) explains, *"a regulation leads to a greater degree of harmonization, since it immediately becomes part of a national legal system, without the need for adoption of separate national legislation; has legal effect independent of national law; and overrides contrary national laws*".

personal data throughout the Union. These differences may 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. This difference in levels of protection is due to the existence of differences in the implementation and application of Directive 95/46/EC.

2.2 Relevance of EU laws protecting privacy and data protection for the research analysis

A fragmented pan-EU data protection environment may not only contribute to legal uncertainty regarding which and how rules apply under national data protection laws, but also whether data protection rules apply in the first place (i.e. whether precursor, legislative conditions are fulfilled as a matter of jurisdictional competence). Promoting more legal certainty in respect of the meaning of the DPD's/GDPR's definitions – and in particular, their definitional criteria against which it may be assessed that personal data exists - is a critical factor for determining the potential application of data protection rules. Put differently, the very interpretation of the key concept of personal data as a scoping provision (gate-keeping term) – which denotes when legal protection through obligations and rights should apply in a processing situation – should also be viewed within the framework of a developing EU regulatory policy to more effectively respect and reconcile data protection law's twin aims.[78]

Hence, in this thesis, an assessment is made regarding which of two personal data approaches – in fixing the point at which data protection rules applies - is better at both:

- **Ensuring that data protection rules apply whenever the rights of individuals could require safeguarding in respect of processing activities** - so that individuals are not inappropriately deprived of a high level of their rights' protection as a result of the model of personal data used. In that respect, unduly restricting the interpretation of the concept of personal data should be avoided, but so should overstretching it to a level of protection that is too high (being unnecessarily so).

- **Promoting the most legal certainty in terms of the application of EU data protection rules** – this means reducing the likelihood of national differences in interpretations of the

[78] That is, the guiding principle of the definition of personal data should be driven by the goal of harmonising the protection of the fundamental rights laid down in EU primary law in respect of processing activities, while also seeking to guarantee the free flow of personal information within the EU, i.e. mutually preserving in an optimal way (thus, maximising capabilities for achieving) both twin aims.

> DPD's/GDPR's concepts of personal data while providing a high level of protection of individuals rights (in particular, their right to privacy) in respect of processing activities.[79]

In effect, which approach is most likely to provide individuals in all MSs with an equivalent high level of legal protection? The analysis also takes into account that any new approach/model must evidence legal coherency with other data protection principles/concepts and the data protection regulatory regime now and in the future.

This assessment is carried out in Chapter 3, first with respect to the DPD regime, then to the GDPR regime, in considering their definitions of personal data, and the importance of the identificatory requirement within such definitions.

2.3 The different elements of the legal definition of personal data

Before starting analysis of the identificatory-approach to the personal data concept, it is useful first to discuss the main elements of its legislative definition under the DPD, followed by an example of its interpretation in one EU MS's national implementation (under the UK's Data Protection Act 1998, DPA). Also mentioned briefly is the GDPR in readiness for a deeper analysis in Chapter 3.

2.3.1 Under the DPD

Per Chapter 1, the DPD defines personal data in its Article 2(a):

> Any information relating to an identified or identifiable natural person ("data subject"); 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. (emphasis added)

As highlighted, four main elements comprise this legislative definition. In a 2007 opinion on the concept of personal data (hereafter, 'WP136'),[80] the WP discussed each of these so-called "building blocks" in interpreting their meaning under EU law, as summarised below:

- **"Any information"** – This requirement relates to the factual matter subject to the DPD when processed. WP136 opines that this sub-concept should be drawn widely, regardless

[79] Bearing in mind that a lack of legal certainty over where the 'legal line in the sand' is drawn can also have a chilling effect on processing activities because of the consequences of getting the decision wrong, not least because of the fact that the full gamut of data protection rules applies when that line is crossed.
[80] Article 29 Working Party (2007, WP136).

of nature, content, or technical format in which it is presented. Both true and false data, as well as viewpoints, are covered.

- **"Natural person" –** This requirement relates to the type of subject to be protected under the DPD, specifically living humans. WP136 also discusses issues related to data about deceased persons, unborn children, and legal persons.

- **"Relating to"** – This requirement refers to the ways in which data must (at least in one way) relate to its subject. WP136 distinguishes three alternative factors, either one of which should be present to satisfy this criterion: '**content'** (the information is given *about* a particular person);[81] '**purpose'** (the information is used, or is likely to be used, to evaluate, treat in a certain way, or influence the status or behaviour of an individual);[82] or, '**result'** (the use of the data is likely to have an impact on a certain person's rights and interests).[83]

- **"Identified or identifiable" –** As discussed above and for in-depth analysis in the next chapter, this identificatory requirement requires that a natural person must be identified or identifiable from the data being processed. WP136 interprets the conditions pursuant to which it opines that it may be considered satisfied.

Additionally, under the DPD, a sub-category of personal data is ring-fenced. Sensitive personal data is defined as personal data revealing: racial or ethnic origin; political opinions; religious and philosophical beliefs; trade union membership; or, the processing of data concerning health or sex life.[84] Special rules apply to its processing. For example, the DPD introduced rules governing the processing of sensitive data, such as that if data subject consent is being relied upon to justify its processing it must be obtained explicitly. Sensitive personal data is discussed further in subsequent chapters.

2.3.2 Under the DPA

Section 1(1) DPA defines personal data as:

[81] WP136 gives two examples of such information: the results of a medical analysis (which relate to the patient), and the information contained in an RFID tag (which relates to the holder of the identity document in which it resides).
[82] For example, WP136 gives the example that a call log of a telephone call made inside a company office can be used to provide information about the maker and the recipient of the call (e.g., to check what time cleaning staff leave their workplace if they are supposed to confirm by phone at what time they lock the premises).
[83] For example, WP136 gives the example that GPS information can be used by a taxi company primarily to improve waiting times and fuel efficiency, but it can also be used by the company to monitor the performance of taxi drivers, whether they respect the speed limits, seek appropriate itineraries, etc.
[84] Article 8(1) DPD.

> [D]ata which relate to a living individual who can be identified - (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

While this definition substantially accords with the DPD's definition, some material differences may be noted in respect of the linguistic choices for the four basic requirements described above, and how these elements have been interpreted under UK law. Most relevant to this analysis are the DPA's interpretation of the third and fourth elements identified above. While the fourth (identificatory) requirement is discussed at length in the next chapter, also worthy of consideration is how UK law has approached the third requirements – i.e. relevant data must also 'relate to' an individual - because of the close nexus between these two elements and their interpretations.

For this reason, the next sub-section summaries what it means for data to 'relate to' an individual under UK law in interpretation of the DPA, in comparison with how WP136 treats this issue under the DPD. This brief review justifies why the focus of this thesis is on interpretation of the identificatory requirement and the traditionally-perceived importance of that element's role in determining the outer boundaries of the personal data concept (by placing a material, theoretical limitation upon the application of data protection law). It also illustrates the tendency for variant approaches to be taken in defining personal data whereby some MSs favour a narrow interpretational approach, while others prefer a broad approach.

2.3.3 Reconciling EU and UK interpretations of the 'relating to' element

For background, in 2003, in what was considered to be a landmark decision of its time, the English Court of Appeal gave a narrow interpretation of the meaning of personal data in the *Durant* judgement. [85] Auld LJ established a combined two-part test for determining whether information relates to an individual:

- whether the information is biographical in a significant sense (it must go beyond the recording of the individual's involvement in a matter or event in respect of which their privacy could not be said to be compromised); and
- whether the information has the putative data subject as its focus.[86]

[85] Durant v Financial Services Authority [2003] EWCA Civ 1746 (8 December 2003).
[86] The key section from the judgment (by Lord Justice (LJ) Auld) is at para.28: "*not all information retrieved from a computer search against an individual's name or unique identifier is personal data . . . Whether it does so in any*

This part of the *Durant* judgement narrowed the interpretation of the concept of personal data under UK law, compared to previous interpretations, by focusing on the scope of the concept in relation to an individual as being about them in a way that affects their privacy.

In 2004, the European Commission sent a letter of formal notice to the UK Government about the conformity of several aspects of the DPA's implementation of the DPD and its application by UK courts, including the implications of the *Durant* case under the DPA. Although the exact contents of that letter have not been made public, apparently the Commission said that it worried that the ruling was inconsistent with the DPD - and its broad construction of personal data - and could compromise the protection afforded to UK citizens.[87] It maintained these concerns until 2010. [88]

In the meantime, in 2009 the ICO issued a technical guidance note (TGN) on what constitutes personal data for the purposes of the DPA.[89] The ICO sought to reconcile the *Durant* judgement with WP136 by offering a wide interpretation of the test for establishing that data relates to an individual: broadly, the processing of the relevant data must be to learn or record something about that individual, or it could have an impact on that individual. Specifically, the TGN states that the following sub-questions should be considered and, if any are answered in the affirmative, the data will be deemed 'relating to' an individual:

- Is the data obviously about a particular individual?
- Is the data "linked to" an individual so that it provides particular information about that individual?
- Is the data used, or is it to be used, to inform or influence actions or decisions affecting an identifiable individual?

---

particular instance depends on where it falls in a continuum of relevance or proximity to the data subject as distinct say, from transactions or matters in which he may have been involved to a greater or lesser degree. It seems to me that there are two notions that may be of assistance. The first is whether the information is biographical in a significant sense…The second is one of focus. The information should have the putative data subject as its focus rather than some other person with whom he may have been involved or some transaction or event in which he may have figured or have had an interest... In short, it is information that affects his privacy, whether in his personal or family life, business or professional capacity".

[87] In a letter dated 9 July 2004, the European Commission wrote to the UK government concerning prospective infraction proceedings against it due to what the Commission considered to be deficiencies in the UK's transposition of the DPD in national law under the DPA. The UK government formally responded in a letter dated 17 November 2005. On 4 April 2006, the Commission again wrote to the UK government explaining its concerns about the UK's implementation of the DPD. See references to these facts in: EA/2012/0110 Ministry of Justice v the Information Commissioner & Dr C Pounder.

[88] In July 2010, after having sent the UK a letter of formal notice, the European Commission sent a reasoned opinion to the UK government in which it criticised the fact that the DPA does not properly implement certain provisions contained in the DPD. See, European Commission (2010, Data protection: Commission requests UK to strengthen powers of national data protection authority, as required by EU law, [online]).

[89] ICO (2007). The format of the TGN is a flow chart of numbered questions to assist those who wish to make determinations over whether data are personal under the DPA, particularly in circumstances where it is not obvious, accompanied by detailed advice and examples.

- Does the data have any biographical significance in relation to the individual?
- Does the data focus or concentrate on the individual as its central theme rather than on some other person, or some object, transaction, or event?
- Does the data impact or have the potential to impact on an individual, whether in a personal, family, business or professional capacity?

Therefore, five parts of these eight-part questions posed by the ICO correlate to the issue of whether the data relates to an individual by content, purpose, or result, the three WP-identified factors. Moreover, as these questions are formulated in the alternative (not cumulatively), the TGN's interpretation of the 'relating to' element is as broad as the WP's interpretation of the same in WP136. The ICO also goes further, however, than the WP in asking two further alternative/non-cumulative questions (on biographical significance and focus) with no equivalents in WP136, being derived from the *Durant* judgement. Said otherwise, the ICO appears to have constructed its guidance on the basis that the latter's two-part test is a helpful consideration, not excluding other possible helpful considerations for determining whether data relates to a person (including scenarios where the data subject's privacy is not affected by the 'relating to' data nexus).[90]

In 2013, the English High Court confirmed the ICO's approach in the TGN in a decision addressing the nature of personal data under the DPA in the context of an exemption from disclosure of such data under the UK Freedom of Information Act (FOIA) 2000.[91] The Court held that, although *Durant* is the leading UK authority on the meaning of personal data, it is limited to a particular factual scenario. It is only one of a number of tests (alongside those set out in WP136 and the TGN) that may be applied in determining whether information relates to an individual and, therefore, could be deemed personal data if they are also identified/identifiable from that data. In 2014, the Court of Appeal followed suit in another case dealing with the data protection exemption contained under FOIA rules.[92]

[90] As a result, although the TGN is entirely consistent with WP136, it is not consistent with the *Durant* case. That was the conclusion the Information Tribunal reached in its decision in Harcup v. Information Commissioner, Information Tribunal, 5 February 2008, when it found (para.20): "*[w]e have difficulty in reconciling the approach in the Guidance [of the Information Commissioner] with that in Durant*".

[91] R (Kelway) v The Upper Tribunal (Administrative Appeals Chamber) and Northumbria Police and R (Kelway) v Independent Police Complaints Commission ([2013] EWHC 2575 (Admin), 20 August 2013). Under section 40 FOIA, the test for deciding whether personal data can be disclosed is whether disclosure to a member of the public would breach the DPA's data protection principles.

[92] Efifiom Edem v. Information Commissioner and Financial Services Authority [2014] EWCA Civ 92 (7 February 2014). In particular, the court here found that a First-Tier (Information Rights) Tribunal – in determining whether the names of Financial Service Authority employees were personal data - had been wrong solely to follow the approach taken in the *Durant* case. Instead, the court specifically referred to the TGN.

In conclusion, the 'relating to' element – as one component to be satisfied in order for data to be deemed personal data - is now applied very widely in EU and UK law so, in effect, all data associated with an individual is very easily caught by the application of this element. Therefore, while the 'relating to' element can be interpreted as intertwining closely with the identificatory requirement, it is the latter element which can be attributed primary limitational functionality in scoping the boundary of the personal data concept and thus the application of data protection law (e.g. when it comes to new types of identifiers and emerging technologies). Indeed, although this argument is not pursued in this thesis, one might argue that whenever someone is identified/identifiable from a piece of information, then it may also be deemed to concern them in satisfaction of the 'relating to' element (i.e., because its content is *about* them).[93]

2.3.4 Under the GDPR

The GDPR (Article 4(1)) defines personal data as, "*any information relating to an identified or identifiable natural person 'data subject'*". Thus, the GDPR retains broadly usage of the four main elements in the DPD's definition of personal data: 'any information'; 'natural person'; 'relating to'; and, 'identified or identifiable' (albeit it is discussed in Chapter 3 that guidance to the interpretation of the latter requirement in the GDPR's Recitals, for example, has evolved from the position under the DPD).

2.4 Chapter conclusion

This foregoing discussion illustrates how what constitutes personal data is shaped by formal legal definitions, but the amorphous nature of those definitions allow for variant interpretations in practice. Specifically, national DPAs and courts have taken different approaches to interpreting definitions, including in respect of the four main definitional requirements (and the criteria for assessing whether such elements exist in a particular data situation) for determining whether data is personal and, therefore, its processing is subject to data protection rules.

It has also been argued that the 'relating to' element does not have a decisive impact on the outer scope of the application of data protection in providing legal classification of personal data. This is because, in the EU (with UK law used in example), it is now generally agreed that this element should be very widely interpreted. Furthermore, to hold that all information that could relate to an

[93] To note, this is a research proposition that has not been explored yet. Nor has an in-depth examination of the 'relating to' element been carried out by the author being beyond the scope of this thesis. However, there is some prima facie support for this argument in guidance from the WP itself (WP136, p.10): "*[i]n the context of discussions on the data protection issues raised by RFID tags, the Working Party noted that "data relates to an individuals if it refers to the identity…of an individual*" (referring, in turn, to Article 29 Working Party (2005, WP105, p.8)).

individual in this broad sense is personal data runs the risk of making all data personal by simply concerning an individual and falling within the ambit of data protection rules. Said otherwise, potentially any data that can conceivably be linked to an individual (in whatever way, by whoever) would be regarded as personal. Thus, we can perceive the logic behind a limiting factor (additional element) to the legal definition; the information must relate to not just its subject, but also that subject must be identified or identifiable. Therefore, the interpretation of the identificatory requirement may be seen as having a decisive impact in transforming data into personal data by way of providing a materiality threshold.

The next research step is to looks at the substance of the identificatory requirement, in terms of what exactly EU and UK law-makers mean by the term 'identified' and 'identifiable', and how these terms have been interpreted by EU and UK policy-makers (notably, the WP and the ICO). The existing legal and policy framework in the UK/EU is considered in relation to this issue, followed by consideration of the GDPR, to extract the theoretical identification-dependent framework that informs a legal understanding of the term personal data.

As mentioned, this analysis recognises that the guiding principle of the definition of personal data should be driven by the goal of harmonising the protection of the fundamental rights laid down in EU primary law in respect of processing activities, while also seeking to guarantee the free flow of personal data within the EU. To such ends, an identificatory-approach to personal data is examined in light of the research questions to assess: whether it promotes a high level of rights' protection to individuals in situations when personal data so determined by the model would be processed; and, whether it also promotes legal certainty (in order to facilitate the free flow of personal data). An alternative approach underpinning the personal data concept may then more easily be compared assessed against these same benchmarks.

Chapter 3 - The identificatory-approach to the legal concept of personal data

Discussions in Chapters 1-2 lead, logically, to questions about how to interpret legally the concepts of 'identified' and 'identifiable' (together 'the identificatory requirement') – together encapsulating an identificatory-approach - in determining when data relating to a living individual are personal under data protection law. This chapter explores this issue in attempting to isolate the nature of this data-person conceptual association, and starts addressing sub-research question 1:

> **Under the existing identificatory-approach to personal data (exemplified by the statement, 'information can only be personal data if it is capable of identifying the individual to whom it relates'), how effective is this approach in terms of realising and reconciling the twin goals of facilitating the free flow of personal data, and safeguarding individual rights?**

It will be seen that, while it is not entirely clear what EU lawmakers mean by "identified or identifiable", the way these terms are interpreted can have significant implications for when information is, or is not, classed as personal data.

In terms of chapter structure:

- **Section 3.1** examines the identificatory requirement under the DPD/DPA.
- **Section 3.2** explores interpretations of the identificatory requirement pursuant to their legislative definitions.
- **Section 3.3** reviews different possible identificatory models of personal data gleaned from the previous two sections, considering both their weaknesses and strengths in terms of their compatibility with data protection's twin aims and exploring the extent to which these aims have been achieved. Additionally considered is the extent to which the GDPR affects this analysis.
- **Section 3.4** concludes.

3.1 The legislative identificatory requirement

3.1.1 The DPD

Article 2(a) DPD describes personal data as meaning, "*any information relating to an identified or identifiable natural person ('data subject')*". It elucidates:

> [A]n 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.[94]

Thus, the DPD makes key distinctions between - not only individuals already "identified", and those merely "identifiable", from data but also – different ways in which identification from data relating to a person is possible. These ways are "directly" (suggesting a data context already possessing sufficient information to enable identification), or "indirectly" (suggesting a context that does not).

The DPD also acknowledges identifiability to be a capability/process encompassing identification opportunities via multiple person-specific factors. The state of identifiability from data is achievable through linking (associating and combining) the relevant data with other pieces of information about the same individual, which individually might not be sufficient to enable identification. This possibility depends upon considering, first, whether sufficient information is already present to enable identification from a single identifier ("*an identification number*"), or person-specific factor.

While Article 2(a) suggests a wide personal data legal definition scope, Recital 26 DPD narrows its ambit:

> [T]o 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.

This qualification (hereafter **'the Means Test'**) has a useful function. It demarcates a 'line in the sand' in the continuum of relevance/proximity between an individual and data relating to them and potentially identifiable of them (the processing of which is deemed worthy of data protection) legally. Effectively, it is a materiality standard/benchmark. Assessing identifiability under the DPD

[94]To note, this definition is also similar to the definition of personal data found in Convention 108 and in the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980, rev.2013).

should consider the possible *means* for effecting identification only where they are *likely reasonably* to be used.[95]

Disagreement exists, however, over the legal status and effect of EU directive recitals. Whereas directive articles give it operative legal force, its preamble-recitals are conceived traditionally as explanatory only and not legally binding.[96] Therefore, taking a risk-averse position, Recital 26 DPD provides points for consideration only (of which account may be taken in deciding whether data are personal or not under EU data protection law); however, the essence of the identificatory analysis to be carried out must follow the (wider) language of Article 2(a) itself. Said otherwise, there is some legal authority for saying that the Means Test cannot cut down the latter's expansive interpretation of when the identificatory element may be met.[97]

Notwithstanding, the reference to "likely reasonably" "means" to achieve identification still implies a relatively broad scope of determination over data types that might be considered personal. This breadth, along with the indeterminacy of its exact scope and its authority, is discussed further below, and in Chapter 5 regarding anonymisation.

3.1.2 The DPA

As mentioned, the DPA defines personal data at section 1(1):

> [D]ata which relate to a living individual who can be identified— (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

---

[95] Without the Means Test, logic would dictate that any information relating to a (living) person is personal data where it contains attributes from which a person may be identified (i.e., including where additional information can be obtained allowing the identification of the data subject in tandem with such information). As the latter scenario would reach very broadly in scope, it can be argued that the Article 2(a) standard is less than helpful in determining the outer remit of data protection law in practice.

[96] For example, the CJEU has declared that directive recitals "*cannot be relied on as a ground for derogating from the actual provisions of the act in question*" (Case C-162/97, Nilsson and Others [1998] ECR I-7477, para 54). It has also been stated that it is a general principle of EU law that the terms of a recital cannot be used to give a narrow construction to the substantive provisions of a measure that its wording would not otherwise bear (Case C-412/93, Societe d'Importatation Édouard Leclerc-Siplec v. TFI Publicité SA and M6 publicité SA, [1995] ECR 179, paras 45-47). Ultimately, therefore, if a recital is (deemed) irreconcilable with an article of a directive, there is authority to argue that the former should be ignored.

[97] See, e.g., the discussion in Google Inc. v. Judith Vidal-Hall and others [2015] EWCA Civ 311 (27 March 2015), at para. 125: *"the starting point must be the wording of Article 2(a) itself…recital (26) should be given an expansive interpretation in the light of the purpose of the Directive as a whole, which is to provide a high level of protection to the right of privacy in respect of the management of personal data by data controllers. To the extent therefore that Article 2(a) and the recital are inconsistent, we think it arguable that…the (wider) language of the provision must prevail."*

Thus, data are personal if they enable a living individual to be identified, either alone (**the section 1(1)(a) test**) or combined with other information (**the section 1(1)(b) test**). As expected, the DPD's/DPA's definitions are similar. Specifically, while there is no DPA mention of the term 'identifiable', the definitional relevance of the notion of data-identifiability is implied by the premise that a person may be deemed capable of being identified from relevant data in conjunction with additional information which (while not currently in the possession of the controller) is *likely* to be sometime.

However, the differences between the DPD/DPA definitions are noteworthy. Strikingly, the DPA adopts a data controller-centric approach. Yet, the section 1(1)(b) test – in focusing on the probability of certain other information coming into the possession of the data controller – has given rise to legal uncertainty. Specifically, it is unclear whether the one capable of effecting an identification under this test is meant to be restricted to data controllers, or whether it could extend to third parties also, i.e. distinct from issues of possession (as explained further below).

Furthermore, the DPA does not refer to Recital 26 DPD. Thus, the reasonable likelihood of possible means being used to identify a person from the relevant data (alone or combination with other information) is a factor for consideration by UK data controllers, but not a statutory requirement.

3.2 Interpretive guidance to the identificatory requirement

There are interpretive gaps to be filled regarding the identificatory requirement so defined in legislation, especially regarding the notion of identifiability associated with contexts where information is insufficient to enable identification from relevant data relating to persons alone. Secondary sources of the type described in Chapter 2 provide much-needed guidance, primarily publications by the WP, the ICO, and judicial authority (including CJEU judgements, and national case-law, interpreting the DPD/DPA respectively).

The ensuing overview breaks down into three thematic question-headings for analysis. These questions represent three, inter-related aspects regarding interpretation of the identificatory requirement, considerations about which helps determine when data protection law applies in practice:

1. **'Identifiable how' (In what ways can identification from personal data be achieved)?**
2. **'Identifiable by whom' (the importance or otherwise of a perspective-dependent viewpoint for determining whether data are deemed personal)?**

3. 'Identifiable with what likelihood of occurrence' (how sure do you have to be that someone could be identified from data before deeming it personal)?

3.2.1 Identifiable how?

3.2.1.1 The WP

In WP136, the WP aims to summarise "*the situations in which national data protection legislation should be applied, and how it should be applied*",[98] to promote a more uniform understanding of the personal data concept.

WP136 starts from the premise that the DPD text (Article 2(a)) invites a broad interpretation of personal data. Regarding the identificatory requirement, the WP states that the normal way to achieve identification from data is through identifiers (subsequently defined as, "*particular pieces of information…which hold a particularly privileged and close relationship with the particular individual*"[99]). In turn, however, the WP comments that - in considering specific types of identifiers through which individuals may become known – knowing a name may not always be sufficient for identification purposes, unless accompanied with other information about the individual.[100]

Conversely, the WP also acknowledges that you can identify someone without necessarily knowing their name. WP136 interprets identification to encompass the possibility of an individual being distinguished from others:

> In general terms, a natural person can be considered as "identified" when, within a group of person, he or she is "distinguished" from all other members of the group. Accordingly, the natural person is "identifiable" when, although the person has not been identified yet, it is possible to do it.[101]

---

[98] WP136, p.12.
[99] Ibid.
[100] Ibid, p.13: "*[a]ll these **new pieces of information linked to the name** may allow someone to zoom in on the flesh and bone individual, and therefore through the identifiers the original information is associated with a natural person who can be distinguished from other individuals*" (emphasis added).
[101] Ibid, p.12. See also p.14: *"a name may itself not be necessary in all cases to identify an individual. This may happen when other "identifiers" are used to single someone out. Indeed, computerised files registering personal data usually assign a unique identifier to the persons registered, in order to avoid confusion between two persons in the file. Also on the Web, web traffic surveillance tools make it easy to identify the behaviour of a machine and, behind the machine, that of its user. Thus, the individual's personality is pieced together in order to attribute certain decisions to him or her. Without even enquiring about the name and address of the individual it is possible to categorise this person on the basis of socio-economic, psychological, philosophical or other criteria and attribute certain decisions to him or her since the individual's contact point (a computer) no longer necessarily requires the disclosure of his or her identity in the narrow sense. In other words, the possibility of identifying an individual no longer necessarily means the ability to find out his or her name. The definition of personal data reflects that fact."*

The WP appears to suggest here that the mere hypothetical possibility of an individual being singled-out from information is sufficient for them to be deemed identifiable for the WP. However, it goes on to clarify that this would only be the case where this possibility takes into account "*all the means likely reasonably to be used either by the controller or by any other person*".[102] In analysing when an individual should be considered identifiable from data, therefore, the WP approves the language used in the Means Test.

Returning to the singling-out notion and how this may be achieved, the WP refers to online identifiers specifically. Where such identifiers attach to particular individuals, the WP states that there may be an intention to connect them with their 'real-world' names/addresses. Alternatively, however, the intention simply may be to attribute certain decisions to them without making any such efforts to uncover, what the WP calls, their identity "*in the narrow sense*".[103]

It is useful to compare here the WP's later 2011 opinion on geo-location services on smart mobile devices (hereafter, WP185).[104] The WP described inextricable indirect links between mobile devices and their users arising because of the unique device numbers transmitted by the former, which – in combination with calculated device locations - can be personal data because they enable nameless owner-individuals to be singled-out.[105] This analysis places emphasis on unique device identifiers in *combination* with other information permitting singling-out (albeit 'link-strength' may also be enriched through subsequent association with *directly* identifying information). Said otherwise, it suggests that the WP would not deem a unique device address *considered solo* to be personal data where used to track a nameless individual.

This issue aside, the WP's position seems to be that identifiability from data is achievable in a legal sense regardless of knowledge of an individual's 'real-world' identity as long as an identifier

---

[102] Ibid, p.15: "*a mere hypothetical possibility to single out the individual is not enough to consider the person as "identifiable". **If, taking into account "all the means likely reasonably to be used by the controller or any other person", that possibility does not exist or is negligible, the person should not be considered as "identifiable**", and the information would not be considered as "personal data"* (emphasis added). Further statements by the WP on this issue are considered below in sub-section 3.2.3.

[103] Ibid, p.14 (in full): *"computerised files registering personal data usually assign a unique identifier to the persons registered, in order to avoid confusion between two persons in the file. Also on the Web, web traffic surveillance tools make it easy to identify the behaviour of a machine and, behind the machine, that of its user. Thus, the individual's personality is pieced together in order to attribute certain decisions to him or her. Without even enquiring about the name and address of the individual it is possible to categorise this person on the basis of socio-economic, psychological, philosophical or other criteria and attribute certain decisions to him or her since the individual's contact point (a computer) **no longer necessarily requires the disclosure of his or her identity in the narrow sense***" (emphasis added).

[104] Article 29 Working Party (2011, WP185).

[105] For example, says the WP, as device identifiers may be further processed in the context of geolocation services, this allows the location of a particular device to be calculated, especially when evidence from different geolocation infrastructures are combined. In particular, the identification of device users may be achieved upon repeated observations, placing an individual at one or more location points over time, which can gain particular significance if linked to a possible address or work place.

contained therein enables back-tracking to them.[106] We can infer the rationale for the WP's position from its 2015 letters addressed to the three EU institutions (the Commission, the European Parliament, and the Council of the EU) engaged in trilogue negotiations to agree the GDPR final text.[107] In these (content-identical) letters, the WP restates its belief that personal data should be defined expansively under the GDPR, in line with technological evolution, taking account of the fact that people can be singled-out, "*on the basis of identifiers or other information* ***and could subsequently be treated differently***" (emphasis added).[108]

3.2.1.2 The ICO

As mentioned, the TGN is a 2007 technical guidance note on what constitutes personal data under the DPA. Although not legally binding, it is persuasive evidence of the ICO's views on the ways in which identification from data is achievable.

---

[106] In WP136, the reasons that the WP gives to found such concern are interwoven with a discussion about pseudonymisation. While this topic is discussed further in Chapter 5 below (in particular, discussion about the meaning of pseudonymisation, and its new legal definition under the GDPR, can be found at Section 5.1), for now the key analogy to be drawn between indirect identifiability and the concept of pseudonymisation is that the latter denotes personal data from which (at least) direct identifiers have been removed. For example, at p.18 of WP136, the WP states that pseudonymised data (such as key-coded data) may be considered as information about individuals that are indirectly identifiable, and thus personal data, (only) if the pseudonymisation is retraceable so that the individual's identity is potentially discoverable: *"[p]seudonymisation is the process of disguising identities. The aim of such a process is to be able to collect additional data relating to the same individual without having to know his identity. This is particularly relevant in the context of research and statistics…**Retraceably pseudonymised data may be considered as information on individuals which are indirectly identifiable. Indeed, using a pseudonym means that it is possible to backtrack to the individual, so that the individual's identity can be discovered, but then only under predefined circumstances**…Key-coded data are a classical example of pseudonymisation. Information relates to individuals that are earmarked by a code, while the key making the correspondence between the code and the common identifiers of the individuals (like name, date of birth, address) is kept separately"* (emphasis added). Of note, in 2014, the WP adopted an opinion (Article 29 Working Party (2014, WP216) where the WP reiterates its view that if it is possible to trace an individual to certain data - considered solo or in combination with other information - data protection rules still apply. However, in WP216 the WP takes its analysis a step further by dissecting the concept of different types of identification risk in the context of considering the robustness of anonymisation techniques applied to personal data. The three types of identification risk described (being singled-out from data, linking data with additional information, and data inference) are discussed at length in Chapter 5. However, for now, it is interesting to acknowledge an apparent change in the WP's position over time from 2007 to 2014 regarding the ways in which it considers indirect (re-)identification from data can be achieved, beyond *just* the ability to single someone out from data. In other words, for the WP in 2014, data may be deemed personal data and its processing subject to data protection law because one or more of these risks cannot be mitigated sufficiently.

[107] See, e.g. Article 29 Working Party. Letter to Ms Ilze JUHANSONE Ambassador Extraordinary and Plenipotentiary Permanent Representative to the EU, Brussels, 17 June 2015.

[108] Ibid, p. 2: "*[t]o ensure the general objective of maintaining a high-level of protection of personal data is upheld, personal data should be defined in a broad manner in line with technological evolution. The definition of personal data should therefore take into account the situation in which people can be "singled out" on the basis of identifiers or other information and could subsequently be treated differently".* Of note, no reference is made in this quote to the WP's previously-cited caveat that only the means likely reasonably used by the data controller or others (including such singling-out means) should be taken into account (see fn.95), although the WP goes on to say directly afterwards that, "*[t]his definition should also take into account the recent CJEU rulings considering why and to what extent IP addresses and other online identifiers are as a general rule to be considered personal data".* Key rulings in this respect are described in the next sub-section, albeit the CJEU did not make reference to the Means Test in its rulings on the personal data status of IP addresses before 2015.

While the TGN is largely consistent with and draws upon WP136, there are differences in approach to interpreting the identificatory requirement. The ICO states that knowledge of an individual's name, together with some other information, will typically be sufficient to identify them. Like the WP, the ICO also opines that information need not necessarily link to an individual's name or address for it to identify them.[109] It gives examples of identification via verbal description ("*the tall, elderly man with a dachshund who lives at number 15*"), or based on physical characteristics (regarding an individual previously unknown to CCTV system operators).[110] Extending the latter example, the ICO says that if operators track "*a particular individual that they have singled out in some way (perhaps using such physical characteristics) they will be processing 'personal data'*".[111]

It is useful to consider the ICO's justification for equating singling-out with identification. In 2001 guidance, the ICO mentions singling-out in discussing identification online:

> If the information about a particular web user is built up over a period of time, perhaps through the use of tracking technology…there might not be any intention of linking it to a name and address or e-mail address. There might merely be an intention to target that particular user with advertising, or to offer discounts when they re-visit a particular web site, on the basis of the profile built up…. Information may be compiled about a particular web user…without any ability to locate that user in the physical world. The Commissioner takes the view that such information is, nevertheless, personal data. In the context of the on-line world the information that identifies an individual is that which uniquely locates him in that world, by distinguishing him from others.[112]

Nevertheless, in the TGN, the ICO also highlights the fact that whether a particular individual may be deemed identifiable from data may not always be obvious, especially when assessing the possibility of the data becoming associated with other information that – in combination – could

[109] TGN (ICO (2007)), pp.7-8: "*[a]n individual is 'identified' if you have distinguished that individual from other members of a group … Simply because you do not know the name of an individual does not mean you cannot identify that individual. Many of us do not know the names of all our neighbours, but we are still able to identify them … There will be circumstances where the data you hold enables you to identify an individual whose name you do not know and you may never intend to discover*".

[110] Ibid, p.6.

[111] Ibid, p.5.

[112] ICO (2001, p.12). The relevance of purpose as a factor to be assessed in construing whether data are personal data is discussed further in Chapter 4. However, to note, in this quote, the ICO appears to count as relevant the existence of a controller's purpose to target an individual in some way after they are singled-out online, while discounting the relevance of purpose to identify the person in the 'real-world' as a necessary condition for identifiability from data. By comparison, the ICO has also opined (IC 2010, p.9): "*personal data is being processed where information is collected and analysed with the intention of distinguishing one individual from another and to take a particular action in respect of an individual. This can take place even if no obvious identifiers, such as names, email addresses or account numbers, are held".*

identify them. In this context, the ICO gives direct weight to the wording of Recital 26 DPD (over the DPA) and concludes:

> [T]he fact that there is a very slight hypothetical possibility that someone might be able to reconstruct the data in such a way that the data subject is identified is not sufficient to make the individual identifiable for the purposes of the Directive. The person processing the data must consider all the factors at stake.[113]

Thus, the ICO in 2007, like the WP's comments that year, exhibits pragmatism in endorsing assessment via the Means Test in non-clear-cut cases.[114] The ICO also states that data should not be deemed personal just because "*there is one organisation that would be able to make such a link [between data it holds to particular individuals] as long as that organisation will not release information enabling such links to be made and adopts appropriate security*".[115] It continues on the next page:

> [A]s long as the first company have appropriate security in place there is little or no chance that any other person who might have access to the coded records would be able to link an individual by name and or address to a particular record. In such circumstances the chances of an individual suffering detriment are negligible.[116]

Said otherwise, the ICO appears to be trying to reconcile here the different identifiability tests in Recital 26 DPD, and the more restrictive DPA section 1(1)(b) test (whereby the other information that would enable identification is limited to that "*which is in the possession of, or is likely to come into the possession of, the data controller*"). In turn, this reconciliatory attempt may be because, for the ICO, the Means Test makes more sense in the context of considering whether the (original) data holder/controller has access to the additional information that would enable it to identify the person to whom the data relates. That is, the ICO's approach here seems implicitly to discount the possibility that a third party might be able to identify the relevant person from the relevant data –

---

[113] TGN, pp.6-7 (in full): "*[s]ometimes it is not immediately obvious whether an individual can be identified or not... for example, when someone holds information where the names and other identifiers have been removed. In these cases, Recital 26 of the Directive states that, whether or not the individual is nevertheless identifiable will depend on "all the means likely reasonably to be used either by the controller or by any other person to identify the said person. Therefore, the fact that there is a very slight hypothetical possibility that someone might be able to reconstruct the data in such a way that the data subject is identified is not sufficient to make the individual identifiable for the purposes of the Directive. The person processing the data must consider all the factors at stake".*

[114] According to the ICO (ibid, p.27), data controllers must consider: the means available to identify the individual and the extent to which such means are readily available; the likely determination likely associated with wanting to identify the individual; and, the feasibility and cost-effectiveness of available means to identify the individual, bearing in mind any new technology or security developments or changes to the public availability of certain records that might be introduced in the future.

[115] Ibid, p.27

[116] Ibid, p.28

in combination with *unknown additional information* - even if the data is currently in the data holder's possession.[117] Sub-section 3.3.2 discusses this issue further.[118]

3.2.1.3 Judicial guidance

The CJEU has held that personal data encompasses data from which individuals can be identified indirectly. In the 2003 *Linqvist* judgement, it referred to identification from an "internet page" by means other than name "*for instance by giving their telephone number or information regarding their working conditions and hobbies*".[119]

More recently, CJEU cases have considered whether IP addresses are personal data. In the 2011 *Scarlet* case, the CJEU found that IP addresses "*are protected personal data because they allow [internet] users to be precisely identified*" without further explanation.[120] In the 2016 *Breyer* case, the CJEU ruled that dynamic IP addresses held by a website operator are personal data under the DPD in certain circumstances.[121] While further discussed below, the latter judgement is worth brief comment now in light of the fact that the CJEU distinguished its factual focus from previous cases by considering IP addresses of the dynamic kind solely. As dynamic IP addresses have variable connectivity with Internet-connecting devices (see fn.10 above), they are weaker enablers than static IP addresses for identifying device owners (or other same-device users) who access a particular website at a relevant time.[122] Moreover, whereas the *Scarlet* case related to the collection

---

[117] In fact, this would be a misinterpretation of the section 1(1)(b) test under the DPA. Literally read - it does not require that the one doing the identifying necessarily be the data controller, only that the additional information enabling identification from the piece of data be in the possession (or likely will come into the possession of) the data controller.

[118] For further discussion in Chapter 5 – by analogy, a controller having access to both raw data and its anonymised data-format may still be deemed subject to data protection rules in respect of that the raw data, whereas it may be considered a separate debate as to whether a *recipient* of the anonymised data (only) would be deemed to be processing personal data.

[119] Case C101-01, Bodil Lindqvist, [2003] ECR I-12971, para.27.

[120] Case C-70/10, Scarlet Extended SA v. Société belge des auteurs, compositeurs et éditeurs SCRL, [2011] ECR I-11959. See also Joined Cases C-293/12 and C-594/12, Digital Rights Ireland and Seitlinger and Others, [2014] ECLI:EU:C:2014:238) which, like the *Scarlet* case on the facts, involved the CJEU addressing whether people could be identified from IP addresses and associated metadata if held by providers with access to additional data that would make identification possible, confirming in the affirmative (at para 26): "*[t]hose data make it possible, in particular, to know the identity of the person with whom a subscriber or registered user has communicated and by what means, and to identify the time of the communication as well as the place from which that communication took place. They also make it possible to know the frequency of the communications of the subscriber or registered user with certain persons during a given period".* However, the latter judgement did not involve considerations of data protection law.

[121] C-582/14, Breyer v. Bundesrepublik Deutschland, [2016] ECLI:EU:C:2016:779. The facts of the German case leading to the referral of questions concerned Patrick Breyer, a German Pirate Party politician, who took action against the Federal Republic of Germany as the operator of publicly accessible websites on which German public institutions supply topical information. He sought, based on data protection law, a prohibitory injunction against the government because it stored IP addresses (including dynamic IP addresses) associated with visitors to their websites for cybersecurity reasons.

[122] Ibid, para 36: *"[I]t is common ground that that the IP addresses to which the national court refers are 'dynamic' IP addresses, that is to say provisional addresses which are assigned for each internet connection and replaced when subsequent connections are made, and not 'static' IP addresses, which are invariable and allow continuous identification of the device connected to the network*". The CJEU also points out, in the context of discussing the features of IP addresses that may stay the same or change each time there is a new connection to the Internet, (para 16): "*[u]nlike*

and identification of IP addresses of internet users carried out by internet service providers (ISPs), the *Breyer* case related to a different scenario. It focuses upon interpretation of the concept of indirect identifiability related to the availability of additional information, which, in combination with the dynamic IP address data, could enable personal identification vis-à-vis (by) non-ISPs.[123]

A more recent CJEU document addressing the issue of 'identifiable how?' is a July 2017 Opinion issued by Attorney General (AG) Kokott in the *Nowak* case.[124] While this Opinion (like all CJEU AG opinions) is non-binding on the final judgement of the CJEU (due towards the end of 2017), its views are still regarded as highly influential. The main issue under consideration is whether a handwritten examination script is the candidate's personal data within the meaning of Article 2(a) DPD. Two points made by the AG are noteworthy.

First, in respect of the argument that the exam script contains the candidate's biometric data in the form of handwriting, and thus is personal data as it may "*provide indications of the identity of the author of the script*", the AG dismisses this argument. [125] She opines:

> The question whether such a handwriting sample is a suitable means of identifying the writer beyond doubt is of no importance for its classification as personal data. Many other items of personal data are **equally incapable, in isolation, of allowing the identification of individuals**

---

static IP addresses, dynamic IP addresses do not enable a link to be established, through files accessible to the public, between a given computer and the physical connection to the network used by the internet service provider".

[123] This, the CJEU explains, is because "*[t]he use by the EU legislature of the word 'indirectly' suggests that, in order to treat information as personal data, it is not necessary that that information alone allows the data subject to be identified*" (ibid, para. 41). The key issue is described at para.39: *"a dynamic IP address does not constitute information relating to an 'identified natural person', since such an address does not directly reveal the identity of the natural person who owns the computer from which a website was accessed, or that of another person who might use that computer….[I]n order to determine whether…a dynamic IP address constitutes personal data within the meaning of Article 2(a) of Directive 96/45 in relation to an online media services provider, it must be ascertained whether such an IP address, registered by such a provider, may be treated as data relating to an 'identifiable natural person' where the additional data necessary in order to identify the user of a website that the services provider makes accessible to the public are held by that user's internet service provider".* Key to the decision, therefore, was the determination of whether the website operator in question (the German government) had the means to obtain such additional information from ISPs for cybersecurity purposes, which, in turn, could enable them to identify particular individuals accessing its website from the relevant (collected) dynamic IP address data.

[124] C-434/16, Nowak v. Data Protection Commissioner, Opinion of Advocate General Kokott delivered on 20 July 2017. The facts of the Irish case leading to the referral of questions to the CJEU concerned Peter Nowak, a failed candidate in a 2009 exam as part of his traineeship to the Institute of Chartered Accountants Ireland (CAI). Nowak issued a (data) subject access request to the CAI in 2010 in respect of his personal data held by them, including his exam script plus any examiner comments on that script. CAI refused on the basis that it was not his personal data and a subsequent complaint by Nowak to the Irish Data Protection Commissioner (DPC) was refused for the same reason. The DPC also held that Nowak's complaint was "*frivolous and vexatious*", not least because - after having reviewed the information - the DPC had identified no substantive contravention of the relevant Irish data protection legislation implementing the DPD (the Irish Data Protection Act 1998, rev. 2003). Nowak appealed that decision in front of Irish courts, ultimately leading the Irish Supreme Court to refer two questions to the CJEU: *"(1) Is information recorded in/as answers given by a candidate during a professional examination capable of being personal data within the meaning of Data Protection Directive? (2) If the answer to Question 1 is that all or some of such information may be personal data within the meaning of the Directive, what factors are relevant in determining whether in any given case such script is personal data, and what weight should be given to such factors?"*

[125] Ibid, para. 29.

> **beyond doubt**. For that reason, neither is it necessary to determine whether the handwriting should be regarded as biometrical information.[126] (emphasis added)

In other words, the AG suggests that, out of context, a person cannot be identified categorically from a sample of their handwriting. This view is at odds with the possibility of indirect identification under the DPD, and section 1(1) of the relevant domestic data protection legislation at issue here (the Irish Data Protection Act 1988, rev. 2003) defining personal data as, "*data relating to a living individual who is or can be identified either from the data* ***or from the data in conjunction with other information that is in, or is likely to come into, the possession of the data controller***" (emphasis added).[127] Notwithstanding, the AG opines that the fact that a script does not bear an exam candidate's name, but instead contains an identification number or bar code, is sufficient for the existence of personal data to be found from which the data subject may be considered indirectly identifiable. This is irrespective of the fact that, as the AG later points out, "*importance is placed on examiners not finding out the 'real-word' identity of the candidates, in order to exclude conflicts of interest or bias*". [128]

Second, the AG lays emphasis, in concluding that the exam script is personal data, on the contextual circumstances under consideration because the identificatory requirement (or, possibly, the 'relating to' element as it is not completely clear which is being discussed) is met purposively in relation to measuring a person's performance:

> [I]n every case, the aim of an examination — as opposed, for example, to a representative survey — is not to obtain information that is independent of an individual. **Rather, it is intended to identify and record the performance of a particular individual, i.e. the examination candidate**. Every examination aims to determine the strictly personal and individual performance of an examination candidate. [129] (emphasis added)

This purposive interpretation is revisited in Chapter 4.

---

[126] Ibid, para. 30.

[127] This peculiarity of this reasoning is highlighted later on in the Opinion by the fact that the AG says – in justifying her belief that an examiner's corrections on an exam script are also personal data – because (para.61): *"[t]he organisation holding the examination is also able to identify the candidate without difficulty and link him with the corrections once it receives the marked script back from the examiner".* Might not the same possibility also be entertained in relation to the candidate's handwriting on the exam script, if a sample of that handwriting (e.g. the candidate's signature) was already on file with the exam board? And yet it was not mentioned, nor was the Means Test. See similar arguments by the thesis author, Knight (2017, Advocate General Delivers Opinion on Whether Examination Scripts Are Personal Data under Data Protection Law [online]).

[128] Ibid, para. 60.

[129] Ibid, para. 24. The AG also concludes later in her Opinion that any examiner's corrections are personal data of the candidate (and potentially the examiner him/her self) because, "*the purpose of comments is the evaluation of the examination performance and thus they relate indirectly to the examination candidate*" (para. 61), and due to the "*close link between the examination script and any corrections made on it"* (para. 63).

In England, two judgements inform the interpretation of the identificatory requirement under the DPA. First, a Court of Appeal 2014 judgement[130] rebutted the argument that determining identifiability is equivalent to determining whether a person is contactable/traceable in practice:

> This argument misunderstands the concept of an identifiable natural person. If a person might be identified by a combination of the name and the context in which it is used it is nothing to the point that it may be difficult to contact them…That is a different concept from whether the person can in fact be contacted or traced.[131]

A year later, in a hearing to decide whether three England-based users of Safari (Apple's internet browser) could make a claim against US-based Google, the Court of Appeal revisited the requirement and its interpretation under the DPA.[132] The claim argued that web browser-generated information (BGI) might be deemed personal data under the DPA because individual Safari users

[130] Efifiom Edem v. Information Commissioner and Financial Services Authority [2014] EWCA Civ 92 (7 February 2014).
[131] Ibid, para. 14. Although, to note, this case was decided in respect of FOIA 2010, which entitles individuals to be provided with information that is held by public authorities unless an exemption applies, including where an FOIA request is for third-party personal data under section 40(2) FOIA. As mentioned (fn.91 above), that section provides an exemption from disclosure if that disclosure would contravene one of the data protection principles in the DPA.
[132] Google Inc. v. Judith Vidal-Hall and others [2015] EWCA Civ 311, 27 March 2015. For background, the English Court of Appeal decided in that case to uphold an earlier decision of the High Court related to the same claim (in Vidal-Hall and others v Google Inc [2014] EWHC 13 (QB), 16 January 2014). By way of a brief summary of the factual background to the claim, the England-resident individuals were pursuing joint proceedings against Google for circumventing the security setting of their Apple's internet-accessing devices so it could install cookies on their Safari internet browser. Since 2011, all versions of Safari made by Apple had been set by default to block third-party cookies planted by parties other than the owner of the website a user visits. However, Apple implemented a number of specific exceptions so as not to prevent the use of certain popular web user-functionality – such as Facebook's 'like' button – on its browser. Google was said to have exploited one of those exceptions, enabling it to track Safari users' online behaviour when they used various Google searches during a nine-month period on their Apple devices, despite publicly stating that such activity could not be conducted unless Safari users gave their consent. Specifically, the allegations centred upon the effect that Google's 'DoubleClick ID' cookies – collecting information from online users who visit websites on which certain subscribing advertisers are present – were stored on users' devices despite the default privacy settings. The users argued that this allowed Google to collect and ultimately aggregate browser-generated information (BGI) about them – including potentially very sensitive personal information about their interests – without their knowledge or consent and to their detriment. Furthermore, because the cookie value of a DoubleClick ID cookie is unique, Google could have aggregated the information it received from different advertisers about an individual user's visit to non-Google websites to create a very detailed profile of their browsing habits. This practice was contrary to the publicly stated position of Google regarding Safari users. Consequently, the group of users (the claimants) alleged, in respect of their claims for misuse of private information and/or breach of confidence, that their personal dignity, autonomy and integrity were damaged. They sought damages for anxiety and distress, and an account of profits against Google. The damage suffered was claimed to arise, not only be virtue of the fact that their BGI was collected by Google without consent, but also because that data was processed specifically to enable advertising to be targeted at users on their devices. In turn, readers of their device screens could then have inferred connections between characteristics associated with displayed online targeted advertising and project them to the users (rightly or wrongly). A legal analysis of this case by the thesis author is contained at Knight (2015, Court of Appeal upholds landmark judgement against Google arising from its exploitation of Apple's Safari web-browser privacy settings, [online]).

could be identified from – or were identifiable from - it.[133] The Court of Appeal (and the English High Court preceding) had to decide whether the claimants' arguments were not plainly wrong.[134]

First, Google argued that it could not identify a particular browser-user by name and "*identification, for the purposes of the DPA means (only) identification by name*".[135] Responding, both courts held it to be clearly arguable that the BGI was personal data under section 1(1)(a) DPA (citing the DPD, WP136, and the CJEU's *Lindqvist* judgement in support). Specifically, the Court of Appeal deemed it clearly arguable that identifying someone from data encompasses the activity of singling-out – as the BGI so enabled - assuming section 1 DPA is interpreted appropriately in alignment with the DPD's provisions/aims.[136]

Second, under section 1(1)(a) DPA, Google argued that recognising a browser on a device (potentially accessed by multiple users) cannot sensibly be said to identify any one individual. The Court of Appeal stated that a clear counter-argument could be made that the concept of multiple single-device users is outdated, as it is typically possible now to equate an individual device user with the device itself. Even if a device has multiple users, the Court of Appeal commented:

> [I]t is the browsing habits of real individual users which are being recognised and tracked … [Google's] "doubleclick" cookie ascribes a unique ID code, to an individual's browser; once it has been set, the defendant can use this unique ID code to identify each time a user subsequently visits a website, uniquely "picking out" the individual.[137]

---

[133] Ibid, para 115: *"(i) BGI information comprises two relevant elements: (a) detailed browsing histories comprising a number of elements such as the website visited, and dates and times when websites are visited; and (b) information derived from use of the 'doubleclick' cookie, which amounts to a unique identifier, enabling the browsing histories to be linked to an individual device/user; and the defendant to recognise when and where the user is online, so advertisements can be targeted at them, based on an analysis of their browsing history".*

[134] This was because both courts did not need to decide the issues in their judgements. Rather, they just needed to determine whether there were serious legal issues to be tried in relation to the claimants' case (i.e., such that the arguments may be left to be considered at a full trial in the future). In fact, the case settled confidentially before a trial commenced.

[135] Ibid, para.112.

[136] Ibid, p.115: "*[i]f section 1 of the DPA is appropriately defined in line with the provisions and aims of the Directive, identification for the purposes of data protection is about data that 'individuates' the individual, in the sense that they are singled out and distinguished from all others. The BGI singles them out and therefore directly identifies them for the purposes of section 1(1)(a) of the DPA … because it tells the defendant (i) the unique ISP address of the device the user is using i.e. a virtual postal address; (ii) what websites the user is visiting; (iii) when the user is visiting them; (iv) and, if geo location is possible, the location of the user when they are visiting the website; (v) the browser's complete browsing history; (vi) when the user is online undertaking browser activities. The defendant therefore not only knows the user's (virtual) address; it knows when the user is at his or her (virtual) home*". Ibid, para.119: "*[t]he best proof of this is defendant's own business model which is predicated on the potential for the "individuation" of users*"; see also para.124: "*the fact that a data controller might not aggregate the relevant information in practice is immaterial. What matters is whether the defendant has "other information" actually within its possession which it could use to identify the subject of the BGI, regardless of whether it does so or not*".

[137] Ibid, para 120.

3.2.1.4 Critical summary

It seems that the ways in which personal identification may be effected (either as a *fait accompli*, or as a possibility, from data) should be assessed broadly. Specifically, authority for recognising the ability to single-out individuals from data as a possible route to their identification indicates a trend toward broadening data protection law in alignment (at least prima facie, because it is expansive in scope) with its aim of safeguarding individual rights in respect of data relating to them to a high level.

Yet, there are still unresolved questions and legal uncertainty about *how* an individual may be distinguished/distinguishable from others in a dataset, taking into account that it seems inter-linked with the issue of *to what extent* an individual may be deemed distinguished/distinguishable.[138] Thus, difficulties remain in asserting exactly what exactly is meant by the terms 'identified' or 'identifiable' from data already established as relating to a particular individual, apart from pointing to evidence that its determination should be context-driven.[139]

In assessing how far the data protection aim of facilitating free flowing personal data - through promoting interpretational consistency - is achievable under the identificatory-approach, policymakers' interpretations of how indirect identification can be achieved appear to rely heavily on the Means Test. However, in the UK at least, courts have primarily opted instead to look to a straightforward construction of the DPA and its section 1(1)(b). The latter's notion of identifiability is restricted to scenarios where other information necessary for identification *"may come into the possession of*" data controllers (regardless of whether or not means are "likely reasonably" to be used to identify a person from those two data sources). Per Chapter 2, one explanation for this inconsistent interpretation of the DPA's section 1(1) definition of personal data is that it incorrectly implemented Article 2(a) DPD. Said otherwise, the DPA is incompatible with the DPD in this respect.[140]

---

[138] For example, making open a dataset about a group of individuals including data relating to you but also a million other members appears a qualitatively different proposition from making open a dataset where the number of other group members from whom your data entry can be distinguished is just fifty. This is because we are more obscure in a larger crowd. Theories by Hartzog around the benefits of obscurity (including online obscurity as a concept) as a form of modest privacy protection are considered in the next chapter.

[139] In WP 136, the WP emphasises the importance of context to the issue of whether data is personal or not (p.13): *"the extent to which certain identifiers are sufficient to achieve identification is something dependent on the context of the particular situation. A very common family name will not be sufficient to identify someone - i.e. to single someone out - from the whole of a country's population, while it is likely to achieve identification of a pupil in a classroom. …So, the question of whether the individual to whom the information relates is identified or not depends on the circumstances of the case".*

[140] Google Inc. v Judith Vidal-Hall and others [2015] EWCA Civ 311, 27 March 2015, para.110: "*Mr Tomlinson says Article 2(a) of the Directive provides for three routes to identification: i.e. data is personal data (i) which directly identifies a natural person; or (ii) from which they are identifiable (a) directly or (b) indirectly. He submits, therefore, that section 1 of the DPA does not accurately transpose Article 2(a) of the Directive into domestic law".*

Either way, Article 2(a) and Recital 26 DPD may be seen as inconsistent. It can be argued that the wording in the DPD's preamble should not cut down the (wider) language of Article 2(a).[141] Alternatively, Recital 26 may also (in any event) be interpreted expansively too in light of the DPD's aim to provide a high level of rights' protection to individuals. Section 3.3 considers this further.

3.2.2 Identifiable to whom?

The key issue is whether the identificatory requirement should be applied 'objectively' (from one viewpoint considered in a non-relativistic fashion), or 'subjectively' (from potentially many viewpoints considered relative to the one doing the identifying). Specifically, assuming a piece of information is deemed personal (because an individual is considered identified/identifiable from that data by the one holding that data):

- Under an objective approach, the information would be deemed personal data in the hands of a third party (regardless of whether a data subject is identified/identifiable to them). Said otherwise, the law would be interpreted as meaning that specific information is personal for everyone that processes it as long as its subject is identified by/identifiable to the first person (typically, the data controller under consideration). (To note, under an absolutist approach, this premise would be extended to the point that as long as its subject is identified by, or identifiable to, *anyone at all*, the information would be personal data).

- Under a subjective approach, the same data would not necessarily be deemed personal for a third party (it would depend upon whether the subject is identified/identifiable from such data by them). Information is personal only for those for whom its data subject is identified/identifiable. Therefore, the determination of whether data are personal would involve considering the context of each data-processing organisation relative to their position.[142]

The Article 2(a) DPD definition does not appear to preclude, at least in principle, the legal status of data being a relativistic matter. Recital 26 DPD explicitly refers to the standard for identifiability as

---

[141] Ibid, para.125: "*[a]s for the second reason, the starting point must be the wording of Article 2(a) itself. Ms Proops submits its (wider) wording cannot be cut down by the wording of the recital, on the general principle of EU law that the terms of a recital cannot be used to give a narrow construction to the substantive provisions of a measure, which its wording would not otherwise bear: see Societe d'Importation Edouard Leclerc-Siplec v TFI Publicité C-412/93 [1995] ECR 1-179, paras 45-47. In any event, recital (26) should be given an expansive interpretation in the light of the purpose of the Directive as a whole, which is to provide a high level of protection to the right of privacy in respect of the management of personal data by data controllers. To the extent therefore that Article 2(a) and the recital are inconsistent, we think it arguable that, as Ms Proops submits, the (wider) language of the provision must prevail".*

[142] A legal analysis by the thesis author on this choice of interpretational positions in the context of discussing the facts of the *Breyer* case is contained at Knight (2016, Mind the Caveats – CJEU Advocate General opines that Dynamic IP Addresses can be Personal Data … (sometimes), [online]).

taking into account "*all the means likely reasonably to be used either* ***by the controller or by any person to identify the said person****"* (emphasis added). This choice of language suggests that a relativistic approach is possible (i.e., account should be taken of any person's perspective for identifying a data subject on a case-by-case basis). Alternatively, it could also suggest an absolutist approach (i.e., the possibility that any person can identify the data subject could be a determinative factor in finding that certain information always be deemed personal).[143] In both cases, however, the linguistic caveat is that it is (*all of the, but only*) the *means* likely reasonably to be used by an array of people to achieve identification that should be considered. So, Recital 26 DPD emphasises less whose perspective on identification should be considered, and more the range of possibilities (capabilities) for achieving identification.

Regarding the DPA, the section 1(1)(b) test appears prima facie to focus on the identifiability perspective (and capabilities) of data controllers, i.e. discounting the possibility of a third party perspective (and their capabilities). However, the section 1(1)(b) test language does not necessarily suggest a strictly objectivist approach (assuming no mis-implementation of the DPD on this point). There is no subject attached to the verb proposition 'can be identified'. Thus, although identification capabilities of the controller are highlighted implicitly, it is arguably not necessary that they be the one able to identify an individual in respect of whom it processes information for it to be personal (only that the additional information enabling identification from the relevant data be in the possession, or likely to come into the possession of, the data controller), provided *someone* can do so.[144]

We look to secondary guidance for further clarification.

3.2.2.1 The WP

As mentioned, WP136 emphasises the importance of context to determining whether the identificatory requirement is satisfied: "*the question of whether the individual to whom the information relates is identified or not depends on the circumstances of the case"*.[145] The WP might

[143] However, it is unlikely to suggest a strictly objectivist approach that only takes into account the perspective of the data controller in determining identifiability.
[144] Compare the argument introduced in fn.117 above. To extrapolate, if a data controller holds both the relevant data and other additional information giving them the means to identify a person from it, they are identifiable to the data controller. Whereas, if the additional information is merely likely to come into the possession of the data controller, the reality is that they are more likely to be in a position to subsequently identify a person from both, but an alternative scenario cannot be excluded. In the TGN, for example, the ICO highlights consideration of what means are available to identify a particular individual and the extent to which such means are readily available. In this context, the ICO states that the data controller should consider the means that are likely to be used by a determined person with a particular reason to want to identify individuals (i.e. it gives the example of investigative journalists, estranged partners, stalkers, or industrial spies being likely to go to further length than the ordinary 'man on the street', at p.9).
[145] WP 136, p.13.

be suggesting legal identifiability should be interpreted in relativistic fashion, depending upon the perspective of the one asking the question whether information is personal data, under a subjective approach. Notwithstanding, in its 2008 opinion on search engines (hereafter, WP148),[146] the WP's direct tackling of the question 'are IP addresses personal data?' also suggests an objectivist (even absolutist) approach on this point.

WP148 opines, although IP address data in many cases are not directly identifiable of people for search engines, identification might be possible by other parties. These include ISPs,[147] law enforcement and national security authorities, and private parties in some MSs (e.g. copyright holders).[148]

Consequently, the WP argues that search engines should consider IP addresses to be personal data, citing Recital 26 DPD in support.[149] Said otherwise, the WP extrapolates from the fact that third parties may be able to achieve identification of a particular individual from data, to arguing that such data can be personal *for another* even when an individual is not identifiable from the data *by them*. Such argument suggests that in interpreting identifiability, the issue of who has the means for identification is irrelevant (taken to the extreme, this suggests that it only matters that *someone somewhere* has such means likely reasonably to be used). Alternatively, the WP's argument here might be seen as ring-fenced to the contentious topic of the personal data status of IP addresses,

---

[146] Article 29 Working Party (2008, WP148).

[147] ISPs may be able to combine them with subscriber information, typically including records of the names, addresses, and banking details of all their customers.

[148] The latter identification capability reflects the likelihood of a copyright owner of a digital product obtaining the corresponding personal details of a (mis-)user to facilitate civil legal action against them. In other words, a copyright holder may obtain a user's identity from an IP address when pursuing abusers of intellectual property rights.

[149] WP148, p. 8: "*[a]n individual's search history is personal data if the individual to which it relates, is identifiable. Though IP addresses in most cases are not directly identifiable by search engines, identification can be achieved by a third party. Internet access providers hold IP address data. Law enforcement and national security authorities can gain access to these data and in some Member States private parties have gained access also through civil litigation. Thus, in most cases – including cases with dynamic IP address allocation – the necessary data will be available to identify the user(s) of the IP address. The Working Party noted in its WP 136 that "… unless the Internet Service Provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data, to be on the safe side". These considerations will apply equally to search engine operators."* In WP136, the WP stated simply under the heading of 'dynamic IP addresses' that *"[t]he Working Party has considered IP addresses as data relating to an identifiable person"* (p.16). In turn, it referred back to Article 29 Working Party (2000, WP37, p.9): *"[i]n these cases, this means that,* ***with the assistance of the third party responsible for the attribution****, an Internet user (i.e. his/her civil identity: name, address, phone number, etc.) can be identified by reasonable means*" (emphasis added). See also its p.21: "*[i]nternet access providers and managers of local area networks can, using reasonable means, identify Internet users to whom they have attributed IP addresses as they normally systematically "log" in a file the date, time, duration and dynamic IP address given to the Internet user. The same can be said about Internet Service Providers that keep a logbook on the HTTP server.* ***In these cases*** *there is no doubt about the fact that one can talk about personal data in the sense of Article 2 a) of the Directive …)"* (emphasis added). To return to WP148, to note however, the WP also described how search engines collect and process vast amounts of user data, including data gathered by technical means, such as cookies, enabling it to track and correlate all the web searches originating from a single IP address, if these searches are logged. In other words, regardless of other perspectives on identifiability, data subjects may still be identifiable to search engines on the facts, where they leave 'traces' which combined with other information received by the servers can be used to create profiles and thereby directly or indirectly identify these individuals.

borne out by the fact that the WP elsewhere gives solid examples of how identification might happen, but such examples do not show that absolutely anyone could fit the bill.[150]

3.2.2.2 The ICO

In section 5.2 TGN - entitled "*different organisations processing the same data for different purposes*" - the ICO states:

> It is important to remember that the same piece of data may be personal data in one party's hands while it may not be personal data in another party's hands … data may not be personal data in the hands of one data controller… but the same data may be personal data in the hands of another data controller…[151]

This suggests ICO endorsement of a relativistic interpretive position on the identificatory requirement under the DPA, and a context-dependent one (reliant upon consideration of the one doing the identifying, the person under identification, as well as the data-embedded identifier(s), in the circumstances). In this way, part of the ICO's reasons for adopting such an approach may come from its consideration of determining what means are likely reasonably to be used for identification as a subjective exercise (requiring assessment of, at least, the capabilities of the one

---

[150] Notwithstanding, in Chapter 5 below, reference is made to other times where the WP takes an objective stance on determining personal data. See, e.g. WP216, p.9: *"[t]he means likely reasonably to be used to determine whether a person is identifiable" are those to be used "by the controller or by any other person". Thus, it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data. Only if the data controller would aggregate the data to a level where the individual events are no longer identifiable, the resulting dataset can be qualified as anonymous… an effective anonymisation solution* ***prevents all parties from singling out an individual*** *in a dataset, from linking two records within a dataset (or between two separate datasets) and from inferring any information in such dataset".* In other words, the WP suggests here that if data controllers retain raw personal data at source, they also retain the ability to attribute it to the relevant individuals, and thus data derived from that raw data that has gone through an effective anonymisation process would nevertheless remain personal data. On the other hand, it can be argued that the analysis should take into account the context and the role of each recipient. See, e.g. El Emam & Álvarez (2014, p.82): "*[t]he European construction of the definition of personal data should therefore not be absolute and take into account the context and the role of each person: a recipient having access to both the source data and the anonymous data would still be deemed a data controller as opposed to a recipient having access only to the anonymized data".*

[151] However, the rest of the quote talks about this turning on the processing purpose in each particular case (for separate discussion in Chapter 4). See an example provided in the TGN, p.15: "*[a]t New Year celebrations in Trafalgar Square two almost identical photographs of the revellers are taken by two separate photographers and stored in electronic form on computer. The first photographer, a photo journalist, takes a picture of the crowd scene to add to his photo library. The second photographer is a police officer taking photos of the crowd scene to identify potential troublemakers. The data in the electronic image taken by the journalist is unlikely to contain personal data about individuals in the crowd* ***as it is not being processed to learn anything about an identifiable individual. However, the photo taken by the police officer may well contain personal data about individuals as the photo is taken for the purpose of recording the actions of individuals who the police would seek to identify, if there is any trouble, so they can take action against them***" (emphasis added). See also the ICO's comment in the TGN that, "*[a] single piece of data, which is not personal data for one data controller may become personal data when it is passed to another data controller*". Another example is given by the WP relating to a scenario about an estate agent taking a photograph of a high street shop to market the property.

under consideration to achieve – or potentially achieve – successful identification).[152] Yet, the ICO also endorses a 'motivated intruder' test for evaluating identifiability from data.[153] This endorsement suggests that determining identifiability does not always depend on the capabilities of the data holder under consideration, but can include assessment of more *non-holder-specific* factors suggesting a more objective approach to identifiability. [154]

3.2.2.3 Judicial guidance

EU judicial authority on this issue is also equivocal.[155] As mentioned, an area of contention regarding interpretation of the identificatory requirement under EU (and MS domestic) data protection law relates to ongoing speculation about whether IP addresses are personal data. A key debate issue is whether it is legally accurate to say that dynamic IP addresses associated with particular web users are always personal data assuming that such users can always be assumed identifiable at least to the ISP. In the 2016 *Breyer* case, the CJEU considered this point of law.[156] The German Federal Court of Justice referred a question to the CJEU as follows (in English translation):

> Must Article 2(a) [DPD] be interpreted as meaning that an IP address which a service provider stores when his website is accessed already constitutes personal data for the service provider if a third party (an access provider) has the additional knowledge required in order to identify the data subject?[157]

To paraphrase the general point of law raised, should the right legal approach to the determination of personal data be subjective (in this case, considering only the

[152] As discussed in the next section, as well as Chapter 4, the Means Test invites consideration of some subjective factors applied to the facts under consideration. In particular, often assessed under the Means Test is the purpose of the processing of the relevant data under consideration. For example, the subjective factors listed by the ICO as relevant to this determination include: identifying the purpose of the processing intended to be carried out on the data; the advantage expected from such usage (including any advantage associated with wanting to identify the individual about whom the data relates); and, the way that the processing is to be structured.

[153] This is discussed further in sub-section 3.2.3.2 and in Chapter 5.

[154] For example, taking into account any new technology or security developments or changes to the public availability of certain records that might be introduced in the future as referred to by the ICO (see fn.114 above).

[155] In relation to IP addresses, for example, in Case C-70/10, Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL, ECR I-11959, where the CJEU found that IP addresses "*are protected personal data because they allow [internet] users to be precisely identified*" - it may be significant that this comment was made in the context of deciding the legality of certain enforcement strategies for the protection of intellectual property rights. To that end, the CJEU may have been concerned about IP addresses being processed for identification purposes to determine whether subscribers are copyright infringers (e.g., by combining IP addresses with content information relating to the types of copyright work accessed to or shared with by the subscribers of the ISP). However, this argumentation was not made explicit in this judgement and no reference was made to who might be able to make the identification.

[156] C-582/14, Breyer v Bundesrepublik Deutschland, [2016] ECLI:EU:C:2016:779.

[157] Said differently, does the legal classification depend on the additional information being accessible in practice to the controller (the German Government acting as a website operator) and its technical and legal capacity to obtain this additional information? Alternatively, should it depend upon what information is available to a third party (an ISP) and, in particular, its additional knowledge that can enable identification of the user associated with a particular dynamic IP address by virtue of its possession of dynamic IP address account-linking customer details?

perspective/capabilities of the (relevant) controller), or absolutist (suffice that identification from an IP address is possible by a third party for it to be deemed personal data also for the controller)?

The CJEU found that a dynamic IP addresses stored by online media service providers who do not have additional information to identify the online user should be regarded as personal data sometimes. It justifies this finding by taking a relativist approach, by looking at identifiability from the perspective of the data controller under consideration, but annexes this issue to consideration of whether the possibility of combining a dynamic IP address with the additional ISP-held information constitutes a "means likely reasonably to be used to identify" the user by such controller.

The CJEU held, in determining this secondary consideration, only legitimate (contrast illegitimate) means by which the controller could obtain such additional information should be taken into account. Such 'legal' means must also not be practically impossible due to disproportionate efforts in time, cost, and manpower. This implies that determining personal data status needs assessing on a case-by-case basis (e.g. taking into account local law provisions). The CJEU concluded on the facts, *"in the event of cyber attacks legal channels exist so that the online media services provider is able to contact the competent authority, so that the latter can take the steps necessary to obtain that information from the ISP and to bring criminal proceedings"*.[158] Therefore, subject to the final assessment of the German Federal Court of Justice, it surmised that, "*the online media services provider has the means which may likely reasonably be used to identify the data subject, with the assistance of other persons, namely the competent authority and the internet service provider, on the basis of the IP addresses stored"*.[159]

A key issue is whether the approach adopted in this decision should be restricted solely to scenarios involving the processing of (dynamic) IP address, or applies to data generally. For example, when determining whether data are personal, should consideration always be given to the possibility of applying for inspection of files in court proceedings or from other authorities - as well as other legal possibilities - to gain access to relevant additional information of a third party? Furthermore, regarding the extent of a third party's involvement, some questions remain even when considered in the context of this case. Notably, the CJEU considered neither the fact that some ISPs delete IP addresses after a certain period, nor the fact that other ISPs do not store IP addresses at all. Does this mean that personal data may

[158] C-582/14, Breyer v. Bundesrepublik Deutschland, [2016] ECLI:EU:C:2016:779, para. 47.
[159] Ibid, para. 48.

become non-personal data, or indeed that identifiability could be excluded upfront upon deeper investigation of the facts? Various other technical and organisational reasons may also prevent identification where IP addresses are concerned. Consider a household's smart-TV reliant on one static IP address to provide online content but where it is almost certain that more than one household member uses the TV, or an internet café where users are not required to register their names/addressees with the café owners. The CJEU did not address such possibilities and their impact on its finding, but they remain relevant contextual factors for possible consideration.

At the EU MS level, national courts have typically not been willing to engage in a detailed analysis regarding the rationale for ascribing IP addresses with personal data status or not.[160] Moreover, there is a split between national courts as to the ultimate answer reached as to whether IP addresses (even static ones) are personal data, depending on interpretation of the relevant provisions of national data protection law implementing the DPD.[161]

In England, the courts traditionally prefer an approach that considers whether data are personal data from the perspective of the data controller.[162] This line of argumentation was used by Google in the *Vidal-Hall* judgements,[163] where Google conceded that it possessed sources of information which (while currently held separately), upon combination might enable it to identify Safari browser users from the BGI (e.g. because Google holds Gmail account data that it could aggregate with the BGI). Google discounted the possibility of third party's identificatory-knowledge being a relevant factor, arguing that even if the Safari browser users were identifiable to others, it did not matter, as the data controller's perspective is the only one that matters under the DPA.[164]

---

[160] For example, in its judgement in Belgian Commission for the Protection of Privacy v. Facebook Inc., Facebook Belgium SPRL and Facebook Ireland Limited 15/57/C, the Brussels Court of First Instance found that a cookie placed on browsers of online users not registered with Facebook constitutes personal data for the sole reason that they contain a unique identifier without further comment (paras.22-23).

[161] Compare, e.g., a German court decision finding that an IP address should be treated as personal data regardless of who holds it (Regional Court Berlin, [2007] K&R, 532), and an Irish High Court decision that IP addresses do not amount to personal data (EMI & Ors v Eircom Ltd [2010] IEHC 108). For more examples, see the results of a (pre-2010) study into EU MSs' case-law on IP addresses as personal data described in Time.Lex, (2010).

[162] See Chapter 5 and Appendix 3.

[163] Google Inc. v. Judith Vidal-Hall and others [2015] ECWA Civ 311, 27 March 2015, upholding Vidal-Hall and others v. Google Inc [2014] EWHC 13 (QB), 16 January 2014.

[164] Ibid, para.111. See also para.126: "*Mr White founds his case on the wording of section 1(1)(b). He submits that the judge's third route to identification involves combining (i) the tailored advertisements (not the BGI) sent to the claimants' devices, and (ii) the knowledge of third parties that a particular claimant uses a particular device; but that the information in (ii) is not likely to come into the possession of the defendant and cannot therefore fall within the second limb of section 1(1) of the DPA".*

However, a counter-argument put forward directly on the issue of identifiability under the section 1(1)(b) test was considered:[165] the claimants submitted that third party knowledge (that a particular device is linked to a particular user) cannot sensibly be excluded from considerations about identifiability under the DPA.[166] The Courts agreed: third parties viewing the claimants' device screens *might* identify the data subjects (as persons having certain characteristics as inferred from the targeted advertisements)[167] and, precisely for that reason, the BGI *might* be deemed personal data.[168] This suggests that a route to identifiability in defining personal data remains possible in future English case-law *other than* through the perspective of the data controller under the DPA (in relation to existing DPA cases still winding their ways through the courts), and/or in its new incarnation under the Data Protection Act 2018 (upon future application in cases brought under the new legislation).

3.2.2.4 Critical summary

Interpretations on the correct perspective from which identification from data (specifically, identifiability) can be achieved under Article 2(a) DPD (as illustrated by interpretational confusion regarding the application of section 1(1) DPA)) are equivocal.

The adoption of an absolutist approach to identifiability would undeniably broaden significantly the applied scope of data protection's aim of safeguarding individual rights in respect of data relating

---

[165] This argument was one of three (the other two mentioned above) on the basis of which the Court of Appeal decided to find that there was a serious issue to be tried regarding whether the relevant BGI was personal data under the DPA.

[166] Google Inc. v. Judith Vidal-Hall and others [2015] EWCA Civ 311, 27 March 2015, para.127-129: *"Mr Tomlinson submits it is plainly wrong to suggest that the only two ways in which a data subject can be identified (or is identifiable) is from the data itself, or from that data together with other information that is held or is likely to come into the possession of the data controller; and that one can exclude the knowledge of third parties from the equation. Ms Proops also submits that the judge's conclusions were sound. In the present case the BGI is processed by the defendant specifically so as to enable advertising to be targeted at users. The targeted advertising is inevitably revelatory as to the browsing history of a particular individual, and hence their BGI. Thus, a notional third party who had access to the device could effectively link the BGI together with the user with the result that the third party would have access to "privacy intrusive" information about known/identified individuals. That a third party is able to "join the dots" in this way and link the BGI to a known individual shows that the BGI must itself be classified as personal data. Ms Proops says that the defendant cannot avoid this result by arguing that it is not likely itself to come into possession of the specific knowledge enjoyed by the third party (that a particular device is linked to a particular user). The defendant has adopted a business model under which its processing of the BGI results in the targeting of advertising at devices; that targeted advertising itself inherently reveals the BGI; and the BGI in turn relates to particular individuals who can be identified by a notional third party with access to the device. In those circumstances the defendant cannot exclude the third party from the analysis made under section 1(1) of the DPA".*

[167] In reminder, Google collected private information about the claimants' internet usage via their Apple Safari browser by means of cookies (the BGI) without the claimants' knowledge and consent. The private information was aggregated and used to create targeted ads based on the claimants' profile, which were displayed on their computer. The adverts revealed private information and were seen or might have been seen by third party bystanders.

[168] Google Inc. v. Judith Vidal-Hall and others [2015] ECWA Civ 311, 27 March 2015, para.128-129. See also para.133: *"[i]t is apparent that the issues raised here are not clear-cut or straightforward. Given our earlier conclusions that there are serious issues to be tried in relation to the claimants' case under both limbs of section 1(1) of the DPA, it is unnecessary for us to say more than that we are not persuaded that the judge was plainly wrong to have had regard to the potential identification of the claimants by third parties. We think this issue is best left to be determined after the facts have been found, and after full argument at a trial."* As mentioned, the claims were settled out of court precluding that possibility.

to them (if information is personal for any one person, it should also be so deemed for other people). Moreover, such a position is compatible with the possibility that the one person for whom an individual is identifiable from data is someone other than the data controller (under the DPA). Indeed, they could be an unknown third party (at least at the time of assessing whether data are personal).

Yet, the CJEU rejected an absolutist position in *Breyer*. Some academic commentators have also rejected it.[169] The implications of adopting an absolutist position are clear. The assumption that if anyone (anywhere) might be able to identify an individual from data, or data and other information, it should be deemed personal has the consequences that all data relating to persons may be presumed personal data. No hypothetical scenario of identifiability – including accidental identification - can be discounted. The objectivist approach – under which the same data would also be deemed personal data in the hands of a third party by virtue of the fact that it is identified by, or identifiable to, a first-hand party (typically construed strictly to be the primary data controller) – also has consequences that many consider non-desirable.[170]

Many, including the ICO (and following English case-law), prefer a relativistic approach that considers the context of the relevant data holder at issue. It aligns with an emphasis upon the circumstances of the relevant data-environment in determining whether data are personal. Arguably, however, such context-dependent analysis lessens legal certainty. As Zwenne points out in referring to the Means Test, interpretive complexities are heightened because, "*[t]he aspect of relativity stands in relation to the aspect of reasonableness"* and *"[a] reasonable amount of effort for one does not have to be the same for another*".[171] Furthermore, a relativistic approach might be considered to narrow the potential application of individual rights in data in the future post-*Breyer* (albeit, the immediate effect of the judgement was broadening the scope of data protection law by clarifying that – whereas ISP-held dynamic IP addresses must be treated as personal data – this was also the case vis-à-vis certain website operators).[172]

---

[169] See, e.g. El Emam & Álvarez (2014, p.82).
[170] This is discussed further in Chapter 5.
[171] Zwenne (2013, p.4).
[172] In other words, the scope of data protection law reaches new dimensions under the *Breyer* judgement in terms of formal acknowledgement of the scope of personal data. The CJEU's view can be summarised thus: (i) it is not sufficient simply that some third party can identify the individual with the data it holds, (ii) the additional identifying data that is held by third parties is, in relation to a controller, only relevant if the possibility to make use of these data constitutes a "means likely reasonably to be used to identify" the individual which requires (iii) that the identification of the data subject must be legally and practically possible for that party without disproportionate effort in terms of time, cost and manpower. As mentioned, a legal analysis by the thesis author on the theoretical issues of debate raised is contained at Knight (2016, Mind the Caveats – CJEU Advocate General opines that Dynamic IP Addresses can be Personal Data … (sometimes), [online]).

This discussion adds another complexity layer in asserting exactly what the terms 'identified' or 'identifiable' should mean. At a minimum – in assessing how far the data protection aim of facilitating the free flow of personal data through promoting consistency of interpretation of the personal data concept is achievable - inconsistency of DPD-implementation and legal uncertainty persists at MS level regarding the identifiability 'from whose perspective' issue.[173]

3.2.3 Identifiable with what likelihood?

The Means Test suggests, in determining whether a data subject may be considered identifiable under EU law, a standard of likely reasonableness may be appropriate. Such a standard would restrict the concept of identifiability implied by Article 2(a) DPD and, therefore, personal data determinations. While the Means Test is non-binding legally, it is worth further consideration here how it has been interpreted at the EU level, with UK law as an exemplar of its domestic interpretation, in this light.

First, to note, the Means test standard read literally does *not* say that it must be likely reasonably that an individual might be identified from data for it to be personal data, although sometimes it has been interpreted so.[174] It only stipulates that account be taken of all the means likely reasonably to be used to identify the individual (from the data). Said differently, it seems that the (potential) data controller should ascertain each means by which someone could identify the individual and then take into account only those likely reasonably to take place.[175] However, there is confusion

---

[173] Korff (2003). As also pointed out by Hildebrandt & Koops (2007, p.39): "*[t]here appears to be a division among Member States on whether or not to use a relative approach to the concept of personal data in the sense that data are considered personal only for someone who can link the data to an identified individual".*

[174] For example, in an anonymisation context, the Means Test has been interpreted as meaning that if it is not 'likely reasonable' that the data recipient would be able to re-identify anonymised data, then that data would not be considered personal for that recipient. However, technically, this is the wrong test. See, e.g. Baines (2015, The Wrong Test for Anonymisation, [online]). Notwithstanding, in WP216 discussed further in Chapter 5, the WP states that it has "*already clarified that the "means … reasonably to be used" test is suggested by the Directive as a criterion to be applied in order to assess whether the anonymisation process is sufficiently robust**, i.e. whether identification has become "reasonably" impossible***" (at p.8, emphasis added).

[175] Although, to note, even the use of the term 'means' in this formulation could raise queries, e.g. whether it should be equated with *any* method (action/process/system etc.) by which a result is achievable, or should be given a more precise meaning in an identification (legal) context? Moreover, how should it be assessed under the DPD whether identification means are available, as well as 'reasonably likely' to be used? For example, to what extent is it meant to encompass consideration around data availability and access? This is discussed further below. To note, for now, an interesting comparison of issues can be made with discussions around reasonableness raised in determining whether personal information exists under the Australian Privacy Act 1988. While that legislation does not adopt a test akin to the Means Test, the test it adopts does encompass an assessment of reasonableness: is an individual 'reasonably identifiable' from certain information (either because the individual can be identified from information in the possession of an organisation, or from that information and other information the organisation may access without unreasonable difficulty)? One question at issue is the extent to which effort is relevant to determining reasonableness under the Australian test. In Privacy Commissioner v. Telstra Corporation Limited [2017] FCAFC 4 CF, for example, Telstra argued that it would be impossible for an organisation to take into account information that they *are likely to access* in deciding whether information is personal information for the purposes of the Australian Privacy Act. Telstra has stated publicly (Australian LRC, 2008, #1, vol.1, p.303): "*[t]he problem with this approach is that it does not seem to require the information to be actually linked or intended to be linked by an organisation for it to fall within the*

over what exactly the Means Test requires a data controller to do *beyond* this exercise to reach a determination on whether information is identifiable of its subject (discussed further below).

Notwithstanding this confusion, it may still be possible to extract insight from interpretations of the Means Test as to the level of likeliness of identifiability that needs to be achieved for data to be deemed personal. Both the Means Test and the question of identifiable with what likelihood involve assessing likelihoods of identificatory capabilities being actuated on the facts. Regarding the question of identifiable with what likelihood, however, assessing whether a degree of identification risk is sufficiently high suggests assessment of the overall picture of identification probability. That is, it would go beyond taking into account the likelihood of individual means of identification being successful, to consideration of such means assessed collectively to determine the overall probability of success, in addition to potentially other factors. In fact, as seen below, the Means Test has been interpreted as calling for consideration of factors other than the means strictly so defined.

3.2.3.1 The WP

WP136 emphasises multiple factors for assessment in considering whether means are likely reasonably to be used by the data controller, or any other person, for achieving identification. These include cost, intended processing purpose, the way the processing is structured, anticipated advantage for the controller if successful, data subject interests, and any risk of organisational and technical failures.[176] The WP provides an illustration of where practical assessment of such factors 'in the round' might lead to concluding that IP addresses relate to identifiable individuals: where copyright holders collect IP addresses *with the purpose of* identifying computer users in order to prosecute them for violation of intellectual property rights. The WP states, in those circumstances, the copyright holder "*anticipates that the "means likely reasonably to be used" to identify the*

definition. Thus, when an organisation collects information about an individual that does not in itself amount to personal information, it would then be required to investigate what other information about that individual is in the organisation's possession in order to determine whether or not the information is to be treated as personal information, even if it does not, and does not intend to, link those items of information. This would be a mammoth task, particularly for large organisations, and would result in increased compliance costs without any clear additional public benefit". Thus, while it may be technically possible for an organisation to identify individuals from information it potentially has access to (e.g. by linking the information with information held by another related organisation), it may be that it is not practically possible (e.g. because of logistics, or legislation preventing such linkage). In these circumstances, Telstra argues that individuals should not be deemed 'reasonably identifiable'. In this case, the Australian Privacy Commissioner disagreed with Telstra, although it was overruled in the end. For more, see Johnston (2017, Mobiles, metadata and the meaning of 'personal information', [online]).

[176] Ibid, pp.15-16: "*[t]he criterion of "all the means likely reasonably to be used either by the controller or by any other person" should in particular take into account all the factors at stake. The cost of conducting identification is one factor, but not the only one. The intended purpose, the way the processing is structured, the advantage expected by the controller, the interests at stake for the individuals, as well as the risk of organisational dysfunctions (e.g. breaches of confidentiality duties) and technical failures should all be taken into account".*

persons will be available e.g. through the courts appealed to (otherwise the collection of the information makes no sense)".[177]

Reference to the intended processing purpose as relevant to Means Test assessments seems, prima facie, at odds with literal interpretation of the language in Recital 26 DPD.[178] It refers only to *means* likely reasonably for use to identify an individual from data – i.e. the means by which identification could be achieved (and correspondingly such means' viabilities). Therefore, one might strictly interpret this test through a literal meaning of the text to signify that in law processing intent should be considered irrelevant. However, its consideration is further justified by the WP. It states that to say that that individual is not identifiable where the controller's intent is precisely to identify them would be nonsensical.[179] Such argument, in turn, suggests that an assessment of means should not encompass factors relevant to an assessment of likelihood (probability) that an organisation would identify a particular person from data in those circumstances where the controller's processing purpose is to identify them precisely because of this presumption-setting line of reasoning.[180]

Nonetheless, WP136 opines that data would not be personal if, taking into account all the means likely reasonably to be used by the data controller or by any other person, the possibility of singling-out an individual from that data "*does not exist or is negligible*".[181] Conversely, to paraphrase, where

---

[177] Ibid, p.17. The WP comments, *"[o]ne relevant factor, as mentioned before, for assessing "all the means likely reasonably to be used" to identify the persons will in fact be the purpose pursued by the data controller in the data processing. National Data Protection Authorities have been confronted with cases where, on the one hand, the controller argues that only scattered pieces of information are processed, without reference to a name or any other direct identifiers, and advocates that the data should not be considered as personal data and not be subject to the data protection rules. On the other hand, the processing of that information only makes sense if it allows identification of specific individuals and treatment of them in a certain way. In these cases, where the purpose of the processing implies the identification of individuals, it can be assumed that the controller or any other person involved have or will have the means "likely reasonably to be used" to identify the data subject. In fact, to argue that individuals are not identifiable, where the purpose of the processing is precisely to identify them, would be a sheer contradiction in terms. Therefore, the information should be considered as relating to identifiable individuals and the processing should be subject to data protection rules."*

[178] As do two other factors listed by the WP: the advantage expected by the data controller, and, the interests of the data subject.

[179] The WP comments (ibid, p.16): *"where the purpose of the processing implies the identification of individuals, it can be assumed that the controller or any other person involved have or will have the means "likely reasonably to be used" to identify the data subject. In fact, to argue that individuals are not identifiable, where the purpose of the processing is precisely to identify them, would be a sheer contradiction in terms. Therefore, the information should be considered as relating to identifiable individuals and the processing should be subject to data protection rules."* To note, the WP uses this argument specifically in relation to the scenario where a data controller *"argues that only scattered pieces of information are processed, without reference to a name or any other direct identifiers, and advocates that the data should not be considered as personal data".*

[180] This would include in respect of situations where a controller holds two or more pieces of information about someone in separate parts of its organisation (or corporate group), which individually would not be able to achieve identification. Yet it can be argued that it does not follow that the controller would in practice combine such information to identify an otherwise unidentified individual (whatever their intent), if it is not reasonably practical to do so. To this end, the likelihood of the organisation linking information with other personal information accessible to it, in view of its end goal, may also be seen as a relevant consideration (including but not limited to considering whether the organisation would be able to link the information without incurring substantial expenditure). See fn.175 above discussing the issues raised by the Australian *Telstra* case.

[181] WP136, p.15 (in full): "*Recital 26 of the Directive pays particular attention to the term "identifiable" when it reads that "whereas to determine whether a person is identifiable account should be taken of all the means likely reasonably*

an assessment has been carried out considering all the ways in which someone might "likely reasonably" identify an individual from data, and it is concluded that there is a *non-negligible* chance of an individual being singled-out from the data on the facts, then it *would be* considered personal by the WP. Thus, the WP seems to link the Means Test to the question of quantification related to the degree of (re-)identification risk that must be present for data to be deemed personal, even if it also links this question to an interpretation of reasonableness. Expressed differently, the WP appears to conflate assessing identification means available to someone (and the overall effort-expenditure to action such means as a proxy for reasonableness), at least in part, with assessing the likelihood of an identification actually taking place.[182] This conflation appears also borne out by a statement in its 2014 Opinion on Anonymisation Techniques (hereafter 'WP216', discussed further in Chapter 5) that it:[183]

> [H]as therefore already clarified that the "means … reasonably to be used" test is suggested by the Directive as a criterion to be applied in order to assess whether the anonymisation process is sufficiently robust, **i.e. whether identification has become "reasonably" impossible**."[184] (emphasis added)

In respect of the practical application of the Means Test, the WP also clarifies that it is not meant to be a static test. Rather, its application is context-specific (i.e. context-dependent related to the means of identification applied to the relevant set of facts at issue). Moreover, the WP opines in WP136 that data may become personal in the future, even if it is currently impossible to identify anyone from the data with all the means now likely reasonably to be used. Put simply, with changing facts over time, there should be a continuous evaluation of risks in respect of the means reasonably likely used to achieve identification from data (including in light of technological developments).[185]

---

to be used either by the controller or by any other person to identify the said person". This means that a mere hypothetical possibility to single out the individual is not enough to consider the person as "identifiable". If, taking into account "all the means likely reasonably to be used by the controller or any other person", that possibility does not exist or is negligible, the person should not be considered as "identifiable", and the information would not be considered as "personal data".

[182] Albeit consideration of both are sometimes speculative as the WP admits. See, e.g. Article 29 Working Party (2013, WP207), where the WP acknowledges that the current interpretation of the Means Test is spurious in practice when applied to unknown third parties. The WP states (p.17): *"[w]hile it is useful to think through the potential motivations of …potential intruders, the WP29 emphasises that there are also considerable limits to this approach: The exercise may be to some degree speculative. In the absence of obvious 'motivating factors' such as those described above the exercise may lead to false reassurances…*" This is discussed further in Chapter 5.

[183] Article 29 Working Party (2014, WP216).

[184] Ibid, p.8.

[185] WP136, page 15: "*[o]n the other hand, this test is a dynamic one and should consider the state of the art in technology at the time of the processing and the possibilities for development during the period for which the data will be processed. Identification may not be possible today with all the means likely reasonably to be used today. If the data are intended to be stored for one month, identification may not be anticipated to be possible during the "lifetime" of the information, and they should not be considered as personal data. However, it they are intended to be kept for 10 years, the controller should consider the possibility of identification that may occur also in the ninth year of their lifetime, and which may make them personal data at that moment."* It goes on: "*the system should be able to adapt to these developments as they happen, and to incorporate then the appropriate technical and organisational measures in due*

Notwithstanding, to return to the IP address example, the WP's guidance is more equivocal. WP136 states that - at least for ISPs and search engines - IP addresses should always be considered personal data to be "*on the safe side*", unless they are in a position "*to distinguish with absolute certainty that the data correspond to users that cannot be identified*".[186] As Zwenne comments, it appears "*no longer important to the [WP] that it may take an unreasonable amount of effort to identify the user"*.[187] While some academics consider this 'absolutist-esque' position by the WP to be disproportionate,[188] the latter's statement does underline a valuable point in relation to questions raised around identifiability with what likelihood. The assessment of the probability of identification occurring is affected by the possibility of false negatives. Hence, it may not be that the WP are saying that IP addresses are always (de facto) personal data as matter of law; rather, its motivation is pragmatic. It wants data controllers to assume such information to be personal ('to be on the safe side'), unless they are positive that it is not, otherwise IP data may be processed unfairly with the result that data subjects may suffer harm (if data protection rules are *not* applied).[189]

---

course". One question left unaddressed by the WP's comments here is whether this on-going assessment of identification risk should include consideration of subsequent, possible availability of additional information about which it later becomes known that identification may be possible in combination with the relevant data. In other words, to what extent does the Means Test require on-going consideration of changes to the relevant data environment beyond technical developments?

[186] WP136, p.17 (in full): "*some sorts of IP addresses …under certain circumstances indeed do not allow identification of the user, for various technical and organizational reasons. One example could be the IP addresses attributed to a computer in an internet café, where no identification of the customers is requested. It could be argued that the data collected on the use of computer X during a certain timeframe does not allow identification of the user with reasonable means, and therefore it is not personal data. However, it should be noted that the Internet Service Providers will most probably not know either whether the IP address in question is one allowing identification or not, and that they will process the data associated with that IP in the same way as they treat information associated with IP addresses of users that are duly registered and are identifiable. So, unless the Internet Service Provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data, to be on the safe side."*

[187] Zwenne (2013, p.6). In this paragraph, Zwenne is talking about what he calls the "reasonableness criterion" (to quote in full): "*the Working Party seems to limit the interpretation of the aspect of reasonableness. In the case of internet café users, the Working Party feels that IP addresses should be treated as personal data after all, even though internet users cannot be traced without an unreasonable amount of effort…because it is unclear which users can, and which users cannot be identified, the Working Party maintains that IP addresses should always be treated as if they were personal data anyway. It is no longer important to the Working Party that it may take an unreasonable amount of effort to identify the user. The Working Party stated that one can only conclude that certain data such as IP addresses are not personal data when the ISP is in a position to distinguish 'with absolute certainty' that the data correspond to users that cannot be identified".*

[188] See e.g. Schwartz & Solove (2011, p.1883).

[189] See e.g. Article 29 Working Party (2000, WP37, p.11): *"[a]t least for security reasons, Internet Access Providers usually seem to systematically "log" the date, time, duration and dynamic IP address given to the Internet user in a file. As long as it is possible to link the logbook to the IP address of a user, this address has to be considered as personal data".* However, this should be compared with what the WP says later at p.21: *"[a]s has been already mentioned in this paper, Internet Access Providers and Managers of Local Area Networks can, using reasonable means, identify Internet users to whom they have attributed IP addresses as they normally systematically "log" in a file the date, time, duration and dynamic IP address given to the Internet user. The same can be said about Internet Service Providers that keep a logbook on the HTTP server. In these cases there is no doubt about the fact that one can talk about personal data in the sense of Article 2 a) of the directive. In other cases, a third party can get to know the dynamic IP address of a user but not be able to link it to other data concerning this person that would make his/her identification possible. It is obviously easier to identify Internet users who make use of static IP addresses. The possibility exists in many cases, however, of linking the user's IP address to other personal data (which is publicly available or not) that identify him/her, especially if use is made of invisible processing means to collect additional data on the user (for instance, using cookies containing a unique identifier) or modern data mining systems linked to large databases containing personally-identifiable data on*

In 2011,[190] by comparison, the WP says that *"[t]he fact that in some cases the owner of the device [transmitting a unique identifier] currently cannot be identified without unreasonable effort*" should not prevent *"the combination of a MAC address of a WiFi access point with its calculated location"* from being treated as personal data.**[191]** The WP rationalises, *"it is unlikely that the data controller is able to distinguish between those cases where the owner of the WiFi access point is identifiable and those that he/she is not*".[192] Again, therefore, the WP appears to consider factors for factual assessment going beyond means accessibility, and the associated reasonableness of effort expenditure, in identifying the data subject. Indeed, such statements hint that determining identifiability in practice should consider identification likelihoods, in tandem with reasonableness. Moreover, in certain cases, it suggests that data may be deemed presumptively identifiable of its subject – whereas, more generally, that presumption may also apply whenever identification is not a remote possibility, or at least is a reasonable probability[193] - except where such presumptions can be overridden by contrary evidence on the facts.

3.2.3.2 The ICO

The TGN states as follows:

> [T]he fact that there is a very slight hypothetical possibility that someone might be able to reconstruct the data in such a way that the data subject is identified is not sufficient to

---

Internet users. Therefore, ***even if it might not be possible to identify a user in all cases and by all Internet actors from the data processed on the Internet****, this paper works on the basis that that the possibility of identifying the Internet user exists in many cases and that large masses of personal data to which the data protection directives apply are therefore processed on the Internet"* (emphasis added).

[190] Article 29 Working Party (2011, WP185).

[191] Ibid, p.11. For background, the WP is discussing here how easy it is to identify the owner of a device transmitting a unique identifier as dependent on the environment. The WP says on the same page, "*[w]hether it requires an unreasonable effort to identify the owners of the WiFi access points, is strongly influenced by the technical possibilities for the controller or any other person to identify them*". See, Knight (2016, Latest Policy Guidance Published on Data Protection and Location Analytics Data, [online]): "*[f]or example, the less populated the local area where the signal come from, the easier it is to determine a single residence and its likely owner with the aid of tools such as house ownership registries. A simple search engine enquiry may also help determine ownership. The WP also refers specifically here to the tendency of people to disclose the location of their houses or work places online, along with other information that can enrich the likelihood of a link being made between geo-location data, an address, and their identification. Even in areas of more dense population, the WP point to the availability of resources such as signal strength, which can help pinpoint the precise location of a Wi-Fi access point that can then be linked to a particular person...Thus, for the WP, the reasonableness of the effort that may be expended in identifying Wi-Fi access points and their owners, is strongly influenced by the technical possibilities for the controller or any other person to pinpoint them physically and then identify them using additional (e.g. publicly-available) information".* Notwithstanding, in areas which are very densely populated, in circumstances where a potential access point location cannot be isolated, the WP admits it *"is not possible without unreasonable effort to ascertain precisely the individual living in the apartment where the access point is located*" (at p.11).

[192] Ibid, p.11.

[193] What these phrases actually mean, however, when 'unpacked' from a legal perspective is nevertheless debatable. In the US, for example, a 'reasonable probability' has been defined as "*a probability sufficient to undermine confidence in the outcome*" (Dyson v. Dormire, 2009 U.S. Dist. LEXIS 90211 (E.D. Mo. Sept. 1, 2009)), but this is just one definition of several possibilities of legal interpretation.

> make the individual identifiable for the purposes of the [DPD]. The person processing the data must consider all the factors at stake.[194]

Said otherwise, despite Recital 26 DPD not being reflected directly in the DPA, the ICO seems to endorse the Means Test here, while interpreting its application broadly to consider all the factors at stake beyond non-slight hypothetical possibilities of identification being achieved.

Thus, the ICO highlights consideration of what means are available to identify a particular individual and the extent to which such means are readily available.[195] The ICO also comments that controllers should consider the means likely to be used by a person determined to achieve identification (one with a particular reason to want to identify individuals).[196] Accordingly, the ICO promulgates a test based on the existence of a competent, diligent, and motivated 'intruder' with access to (legal and non-specialist) resources commensurate with their motivation. The nature of the data (specifically, how this might inform or influence someone's intentions in processing it) is a factor the ICO says could give rise to their motivation for trying to identify the data subject.[197] The ICO also refers to the cost-effectiveness of means to identify individuals in light of new technological/security developments, or public records-availability changes, occurring over time.

To summarise, for the ICO, the Means Test should be assessed practically in light of both objective and subjective criteria. One objectified aspect – not depending on consideration of one particular perspective – includes the means readily available to identify an individual by a hypothetical intruder (bearing in mind their feasibility and cost-effectiveness). Other objective-type factors include technological state-of-the-art, the public availability of information associated with the relevant data under consideration, and data subject interests. More subjective factors - dependent upon a particular viewpoint – could be said to include processing purpose/intent and the way it is

---

[194] TGN, p. 8. See Chapter 5 and its discussion of re-identification risk as compared with identification risk. To note, the ICO develops its position in 2012, when it commented that the risk of re-identification from data that has been subject to anonymisation techniques must be greater than remote, and reasonably likely (in its Anonymisation Code of Practice, 2012, p.6): "*[t]he DPA does not require anonymisation to be completely risk free – you must be able to mitigate the risk of identification until it is remote. If the risk of identification is reasonably likely the information should be regarded as personal data*". In this context, the ICO appears to be alluding to the wording of Recital 26 DPD (over the DPA).

[195] TGN, p.9: "*[f]or example, if searching a public register or reverse directory would enable the individual to be identified from an address or telephone number, and this resource is likely to be used for this purpose, the address or telephone number data should be considered to be capable of identifying an individual".*

[196] In the TGN, the ICO highlights consideration of what means are available to identify a particular individual and the extent to which such means are readily available. In this context, the ICO states that the data controller should consider the means that are likely to be used by a determined person with a particular reason to want to identify individuals and It gives the example of investigative journalists, estranged partners, stalkers, or industrial spies being likely to go to further length than the ordinary 'man on the street' (at p.9).

[197]Article 29 Working Party (2013, WP207, pp.16-17): "*[f]or example, an intruder – in general - might be more motivated to re-identify personal data if such data: have a significant commercial value …and can thus be bought and sold for financial gain; can be used for law enforcement or intelligence purposes; reveal newsworthy information about public figures; can be used for political or activist purposes…; could be used for bad-intentioned personal reasons…; or, could raise curiosity…".*

to be carried out, the determination likely associated with wanting to identify an individual (including advantage expected), and any disclosure-limiting safeguards put in place.

Part of the ICO's reasons for adopting a more dynamic (contrast static) approach than the WP in this respect may flow from its conception of the Means Test as mostly a subjective exercise (dependent on the position of the one doing, or potentially doing, the identifying). Nevertheless, the TGN does not elucidate on how likely it should be that the data subject will be identified on the facts following this analysis beyond excluding the mere fanciful.

3.2.3.3 Judicial guidance

The CJEU has only mentioned in passing degrees of likelihood that someone may be identified from data when assessing whether personal data exists. In the *Breyer* case, the CJEU stated that the means test would not be satisfied:

> "[I]f the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and manpower, **so that the risk of identification appears in reality to be insignificant**. (emphasis added)[198]

By comparison, in the previously-cited English *Vidal Hall* litigation, Google argued that - as it was not "*reasonably likely*" that Google would aggregate the two sets of data in its possession to identify its users from BGI – they should not be deemed DPA-identifiable from the BGI *"and other information which is the possession of, or is likely to come into the possession of, the data*

[198] C-582/14, Breyer v. Bundesrepublik Deutschland, [2016] ECLI:EU:C:2016:779, para. 46. By comparison, also noteworthy are comments made by the CJEU in its judgement in the *Digital Rights Ireland Ltd* case (Joined Cases C-293/12 and C-594/12, Digital Rights Ireland and Seitlinger and Others, [2014] ECLI:EU:C:2014:238) concerning the retention, by providers of publicly-available electronic communications services or of public communications networks, of certain data which are generated or processed by them. While not commenting directly on the DPD and whether personal data was involved, the CJEU took a broad-brush approach to the nature of the relevant data, at least in terms of assessing its likeliness for giving rise to questions relating to respect for private life and communications, and the protection of personal data (Articles 7 and 8 of the EU Charter, respectively) (paras 26-27): *"the data which providers of publicly available electronic communications services or of public communications networks must retain…include data necessary to trace and identify the source of a communication and its destination, to identify the date, time, duration and type of a communication, to identify users' communication equipment, and to identify the location of mobile communication equipment, data which consist, inter alia, of the name and address of the subscriber or registered user, the calling telephone number, the number called and an IP address for Internet services.* ***Those data make it possible, in particular, to know the identity of the person with whom a subscriber or registered user has communicated and by what means****, and to identify the time of the communication as well as the place from which that communication took place. They also make it possible to know the frequency of the communications of the subscriber or registered user with certain persons during a given period. Those data, taken as a whole, may allow very precise conclusions to be drawn concerning the private lives of the persons whose data has been retained, such as the habits of everyday life, permanent or temporary places of residence, daily or other movements, the activities carried out, the social relationships of those persons and the social environments frequented by them*" (emphasis added). In other words, the quote here suggests that the CJEU relies on a 'possibility' (may occur) standard here without delving deeper, albeit as mentioned the case did not involve consideration of the application of data protection law.

controller".[199] Both the High Court and the Court of Appeal rebutted this argument; it was sufficient that data "can be used" to identify an individual (at least insofar as they agreed that this counter-argument should be tested at full trial).[200]

A more in-depth discussion of the 'identifiability with what likelihood?' issue under the DPA is contained in a 2011 UK Upper Tribunal (Administrative Appeals Chamber) decision.[201] The Ministry of Defence (MoD) argued against the release of requested information about soldiers following a freedom of information request, even though the soldiers' names would be redacted, because it might constitute personal data for third parties. The Tribunal noted that - while the Information Commissioner had referred in argument to whether there was an "*appreciable risk*" of (re-)identification - the statutory test uses the phrase "*can be identified*". [202] However, it concluded, "*[o]n the evidence that we have received, our conclusion on the balance of probabilities is that publication of the information the subject of the MOD's appeal will not render individuals identifiable*".[203] Thus, the standard of proof regarding identifiability applied in practice in this case was 'on the balance of probabilities'.[204]

3.2.3.4 Critical summary

A key question under the identificatory-approach to personal data is the significance to be attached to the probability of identifiability from data. How much identification risk is sufficient in the relevant context under consideration for data to be considered identifiable, i.e. how sure does one have to be that the data subject can be identified from data before one deems it personal data?

---

[199] Google argued that the BGI is not personal data under section 1(1)(b) of the DPA, insofar as it stores the collected BGI separately from other (additional) information held by it from which an individual could be identified (such as Gmail accounts).

[200] See, e.g. Google Inc. v. Judith Vidal-Hall and others [2015] EWCA Civ 311, 27 March 2015, para.124: "*[i]n our view, this cannot be a 'knock-out' point for the defendant for two reasons. First, because of the wording of section 1(1)(b) itself; and secondly, because it raises a substantial issue as to the correct interpretation of Article 2(a) which will not obviously be resolved in the defendant's favour. As regards the wording of section 1(1)(b), this refers simply to information "in the possession of" the data controller, and appears only to be concerned with whether data "can be used" to identify an individual (not with whether it has been used or is intended to be used in this way). On a straightforward and literal construction of the section, therefore, the fact that a data controller might not aggregate the relevant information in practice is immaterial. What matters is whether the defendant has "other information" actually within its possession which it could use to identify the subject of the BGI, regardless of whether it does so or not".*

[201] All Party Parliamentary Group on Extraordinary Rendition v. The Information Commissioner and the Ministry of Defence [2011] UKUT 153 (AAC). See also reference to this case in Appendix 3 of this thesis.

[202] Google Inc. v. Judith Vidal-Hall and others [2015] EWCA Civ 311, 27 March 2015, para.124.

[203] Ibid, para.129

[204] Essentially, the balance of probabilities evidential (also known as the 'preponderance of evidence') standard is linked to a 'more likely than not' test (requiring that it be more likely than not that something occurs/occurred in a certain way). See, e.g., comments by Lord Nicholls **(**Re H (Minors) (Sexual Abuse: Standard of Proof)) [1996] AC 563, para 73)**:** *"[t]he balance of probability standard means that a court is satisfied an event occurred if the court considers that, on the evidence, the occurrence of the event was more likely than not*". Applied to identifiability, this would mean that it would have to be assessed whether the possibility of achieving identification of the person to whom data relates is more likely than not to occur.

For example, "*is mere possibility enough, no matter how distant, or must identification actually, in the instant case, be a realistic possibility?"*[205]

Mostly this issue is side-stepped in interpretive publications about the DPD/DPA personal data definitions. Where comments are made, typically they are indirect ones via extrapolation of advice on interpreting and applying the Means Test practically (including through use of the ICO-endorsed motivated intruder tool, and considerations around reasonable efforts). While, as mentioned, according to a literal interpretation of Recital 26, this should only extend to consideration of the *means* likely to be used by which a person could be identified from data, this has been expanded to consideration of other factors beyond means (and not under the guise of factors to help interpret means either). Yet there is disagreement over the exact factors for consideration in determining whether that standard has been satisfied, to what extent they hold weight, and to what extent such factors should have a subjective element. For example, for the WP and the ICO, identifiability appears to be a contextual concept, with the ultimate decision on whether personal data is involved involving a multi-factorial assessment by data holders on the facts before them. Notwithstanding, both bodies stray into recommending factors that touch on the issue of how likely it is that identification will be achieved.

In general, this approach suggests that the Means Test be interpreted expansively in keeping with the DPD's aim to provide a high level of rights' protection to individuals. Yet, there remain unanswered questions about whether this is the correct approach legally (putting aside questions about the relationship between Article 2(a) and Recital 26 DPD) and how the Means Test should be applied in practice. For example, as mentioned, to what extent should data controllers give weight to the fact that they hold additional information in other parts of their organisation from which a data subject could be identified in conjunction with the relevant data (regardless of whether they have not already aggregated, and do not intend to aggregate, the two data sets in practice)? Said otherwise, there is uncertainty over whether it is existence, or accessibility, of the additional information that is significant.[206]

Partly, this problem stems from the fact that the Means Test is not framed in terms of identification as an overall possibility of success (a likelihood threshold). Moreover, the relationship between probability and reasonableness in construing identifiability in practice is unclear, e.g. should

[205] Booth et al (2004, p.110) (in full): "*[o]ne of the issues that must be addressed by this 'ideal type' is …how significant is the likelihood of the relevant context arising? Is mere possibility enough, no matter how distant, or must identification actually, in the instant case, be a realistic possibility?"*

[206] Ibid, p.123: "*[i]f, in constructing a decision making strategy on this idea type, a country were to adapt it in order to take account of context then it would still need to establish whether it is simply the existence, or the accessibility, of the additional relevant/necessary information that is significant. If it is accessibility that is crucial, then there is also a question of who is able to access it (e.g. the data controller (as in The UK) or absolutely anybody)".*

identification be reasonably likely to occur? Alternatively, should there be, e.g., a realistic prospect of identification occurring assessed reasonably?

3.3 Critical review of the identificatory-approach and its compatibility with the twin aims

This section reflects on the strengths and weaknesses of the identificatory-approach vis-à-vis the twin goals of facilitating the free flow of personal data and safeguarding individual rights to a high level under existing law. It also examines the GDPR's forthcoming impact on this analysis.

3.3.1 Existing law

Many open-ended questions persist regarding how to interpret the identificatory requirement in the DPD/DPA. While not made explicit in the core legislative texts, secondary source interpretations partly acknowledge the existence of such questions, but vary in the answers provided such that much remains unresolved, or new uncertainties and questions arise. Specifically, there appear to be trends in secondary sources towards admission of the following:

- **Encompassed with the concept of what it is to be identified/identifiable is the notion of mere individualisation from others based on a data-identifier**. Without further qualification, this acknowledgement would broaden data protection law's coverage, providing a very high level of protection for the rights of individuals that could require safeguarding in respect of processing activities.[207]

- **Tension between a more objective perspective on identifiability and a subjective/relativistic one**. Identifiability objectification, or its extreme position of absolutism ('if there is the capacity for identification of a person from data for any one person, so there should also be for others'), would also promote very high levels of the protection of the freedoms/rights (especially, regarding privacy) of data subjects. Notwithstanding, a subjective approach has strong legal authority after the *Breyer* judgement although, notably, objective factors have been imported increasingly into the DPD's Means Test.

[207] See, e.g. Article 29 Working Party. Letter to Ms Ilze JUHANSONE Ambassador Extraordinary and Plenipotentiary Permanent Representative to the EU, Brussels, 17 June 2015, p.2: *"[t]o ensure the general objective of maintaining a high-level of protection of personal data is upheld, personal data should be defined in a broad manner in line with technological evolution. The definition of personal data should therefore take into account the situation in which people can be "singled out" on the basis of identifiers or other information and could subsequently be treated differently".*

- **Determining whether someone is identifiable requires assessment of the data-circumstances, including considering the means likely reasonably to be used to identify them, which can change dynamically over time.** Determining whether personal data exists legally involves a context-dependent exercise, implying that is not possible to exclude the premise that personal data are involved in any one case where information relating to a person is processed.[208] Indeed, in some cases (e.g. by the WP in relation to IP addresses) it may be presumed always to be so. Again, this interpretation supports a high level of individual rights protection. Putting aside what the term 'means' may actually mean (an issue not explored in depth here, see fn.175), an expansive interpretation would also provide a high level of rights' protection to individuals in respect of the processing of data relating to them.

Therefore, these three points favour broadening the scope of data protection law, making it more likely that its principles apply where the rights of individuals could require safeguarding upon processing information related to them; and, conversely, that such individuals are not deprived (inappropriately) of their rights' protection because of the model of personal data used.[209]

Yet, in summing up the critical review of the interpretation of the identificatory-approach under existing law, two main challenges remain unaddressed regarding legal certainty linked to these three points.

First, interpretational stretching of the concept of personal data under the DPD (and national implementing laws) increases the likelihood that certainty becomes replaced with unpredictability undermining the efficacy of the data protection regime.[210] Second, the wider the definition of personal data, the more questionable it is to what extent the second objective – the free flow of personal data across borders – can be achieved. This is because if a legal basis – such as consent - has to be found before (e.g.) an IP address can be processed, it is liable to impede related data flows (see fn.60 above).

Even if you ignore this second argument (because it means that the twin aims are inexorably in tension), legal uncertainty over the appropriateness of adopting the three trends highlighted above persists. Thus, a lack of approximation of national data protection laws through differing

[208] Zwenne (2013, p.4).
[209] See, e.g. comments by the WP (WP136, p.5): "*unduly restricting the interpretation of the concept of personal data should … be avoided*". In other words, a restrictive interpretation of the definition of personal data runs the risk of being too restrictive, such that data protection would be under-applied and hence data subjects may suffer if some data relating to them were processed unfairly.
[210] Again, this would be less than optimal and, indeed, could make the data protection regulatory system unworkable.

interpretations poses an obstacle to the flow of personal data that may inhibit it.[211] Additionally, this uncertainty may result in an erratic imposition of strict regulation in relatively innocuous cases (and vice versa).[212] That is, the degree of protection provided by data protection law may be considered disproportionate to the data privacy risks in relation to which data subjects may be deemed vulnerable.[213] Yet, the principle of proportionality is recognised as a key principle of EU law.[214]

Discussion of key associated criticisms in elucidation of this challenge follows, as summarised in the next three sub-headings. This discussion also allows some of the key points raised so far in this chapter to be re-evaluated *across* the three chosen themes (identifiability 'how', 'from whose perspective', and 'with what likelihood').

3.3.1.1 The indeterminacy of the scope of personal data as a legal definition causes practical uncertainty

In determining whether an individual may be deemed identifiable under EU/UK data protection law, the preponderant view appears that those holding data relating to persons should engage in a risk-centric assessment (from which they can then determine whether their planned processing activities are caught by data protection rules). Engaging in such assessments, in turn, depends on considering the data-environment at issue.

For that reason, many would argue that consideration of individual circumstances on a case-by-case basis is required to determine whether an identifier (e.g. a dynamic IP address) constitutes personal data under the identificatory-approach, even if this assessment is carried out according to a definitional framework of analysis to be engaged in.[215]

---

[211] Per Chapter 2, see Recital 8, DPD: "*in order to remove the obstacles to flows of personal data, the level of protection of the rights and freedoms of individuals with regard to the processing of such data must be equivalent in all Member States; … this objective is vital to the internal market but cannot be achieved by the Member States alone, especially in view of the scale of the divergences which currently exist between the relevant laws in the Member States and* ***the need to coordinate the laws of the Member States so as to ensure that the cross-border flow of personal data is regulated in a consistent manner*** *that is in keeping with the objective of the internal market …; … Community action to approximate those laws is therefore needed"* (emphasis added).

[212] This concern is exacerbated by the fact that data protection law draws a bright-line distinction between data that are personal (and subject to data protection rules) and data that are not. See, ibid: "t*he scope of the data protection rules should not be overstretched. An undesirable result would be that of ending up applying data protection rules to situations which were not intended to be covered by those rules and for which they were not designed by the legislator".*

[213] See, also, Schwartz & Solove (2013, #1, pp.913-4): *"[i]t also risks activating burdensome regulations for data-processing entities that are incommensurate with actual risks to the privacy of individuals"*.

[214] See, e.g. Article 5 of the EC Treaty, stating that "*any action by the Community shall not go beyond what is necessary to achieve the objectives of this Treaty*". More specifically to a discussion on privacy/data protection, the Charter (Article 52(1)) states: "*[a]ny limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others*".

[215] See, e.g. Zwenne (2013, pp.4-5): "*[w]hat may, at a given time, in a given situation, for a given individual be deemed to be personal data, may … for another person not be deemed as such".*

The practical uncertainty that this requirement engenders is accentuated by the nature of the assessment tools used to determine identifiability. For example, data controllers have to guess what other information is available to third parties (and, under the DPA, is likely to come into the data controller's possession) which, in combination with the data concerned, might enable individuals to become identifiable.[216] Furthermore, assessments of reasonableness (e.g. what constitutes reasonable efforts to effect identification reliant upon subjective factors, as well as more objective ones) adds to the dynamic status of the personal quality of data that must be determined.

The engendered practical uncertainty in making determinations whether data are personally identifiable of its subject is complicated by the fact that when we drill down into the phrase "means reasonably likely", we find yet more uncertainties:

- The Means Test standard relates to means 'likely reasonably', not that the risk of identification should be "reasonably likely". Arguably, this could require consideration of all of the means likely reasonably to be used for identification purposes, followed by assessment of which of these might be successful and then based on that analysis, coming to a conclusion about the overall likelihood of identification. However, this is not made clear.

- The WP acknowledges that the current interpretation of the Means Test - such as the 'motivated intruder' rule of thumb put forward by the ICO – is speculative in practice when applied to unknown third parties and may, therefore, lead to false conclusions being drawn.[217]

---

[216] Albeit the ICO at least (in its 2012 Anonymisation Code of Practice, pp.24-26) appears to exclude consideration of hypotheticals that are at most merely an educated guess. It states - under the headings "*Prior knowledge and re-identification ...Information, established fact and knowledge*" - *"[t]he starting point for assessing re-identification risk should be recorded information and established fact. It is easier to establish that particular recorded information is available, than to establish that an individual – or group of individuals - has the knowledge necessary to allow re-identification. However, there is no doubt that non-recorded personal knowledge, in combination with anonymised data, can lead to identification. It can be harder though to substantiate or argue convincingly. There must be a plausible and reasonable basis for non-recorded personal knowledge to be considered to present a significant re-identification risk…Identification and the educated guess Data protection law is concerned with information that identifies an individual. This implies a degree of certainty that information is about one person and not another. Identification involves more than making an educated guess that information is about someone; the guess could be wrong. The possibility of making an educated guess about an individual's identity may present a privacy risk but not a data protection one because no personal data has been disclosed to the guesser. Even where a guess based on anonymised data turns out to be correct, this does not mean that a disclosure of personal data has taken place".*

[217] See Chapter 5 for further discussion in the context of assessing re-identification risk, as well as comments by the WP in Article 29 Working Party (2013, WP207, p.17): "*[w]hile it is useful to think through the potential motivations of …potential intruders, the WP29 emphasises that there are also considerable limits to this approach: The exercise may be to some degree speculative. In the absence of obvious 'motivating factors' such as those described above the exercise may lead to false reassurances…*"

- The WP and ICO appear to change their positions on some aspects of identifiability including in relation to scenarios where the Means Test may be used. For example, on the issue of relativity, the WP in 2008 said that, because third parties are considered able to ascertain the identities of particular users, so search engines should consider IP address to be personal data "t*o be on the safe side"*.[218] Yet when presumptions should be used to avert 'unsafe' results is not clear. Indeed, as mentioned, the discussion around whether (and when) IP addresses are personal data (especially dynamic ones, because an online user is identifiable from related data via the assumption that they used an internet-connecting device at a specific point in time) illustrates the inherent 'push-and-pull' of the identifiability interpretive debate.[219] It is unclear to what extent such discussions - and the findings of the *Breyer* judgement – can and should be read across to other identifiers (including other online identifiers specifically, such as browser-enabled cookie files).

3.3.1.2 The open-endedness of the future risk-assessment process compounds indeterminacy practically

An exercise in risk-assessment is necessarily associated with analysing identification possibilities - and speculating about what could happen – indeed, what is likely to happen - in the future, even if conclusions are drawn based upon current 'known knowns'. Yet, trying to predict and quantify identification possibilities can be extremely difficult. Not least, what may now, for a given individual, be deemed personal data, may in another situation at a different time not be so deemed due to a known change (or anticipated change) in external factors, even if the same processing set-up is used.

Per Chapter 1, a major uncertainty factor in assessing identification risk is the increasing availability of data about us in the public domain. Another factor relates to who might get hold of data relating to us in the future. Per Chapter 1 again, while the act of identification may be described as a relational process between data observer and data subject, data-identifiers associated with us are now potentially relational in ways unforeseen with people unknown.[220] This fact can be difficult to encapsulate in a risk-based assessment of identifiability on the facts.

---

[218] Compare comments by the WP in WP136 about how the scope of the data protection rules should not be overstretched (p.5): *"the clarifications in recitals 26 and 27 of the Directive show how the legislator wanted to see data protection applied… Another general limitation for the application of data protection under the Directive would be processing of data under circumstances, where means for identifying the data subject are not "likely reasonably to be used" (recital 26)".*

[219] In other words, it reflects a perception of inherent tension between the twin aims of data protection law.

[220] Said otherwise, in the future, there is no limit on the number of people who may become capable of identifying us in a big data (and increasingly open-access data), hypothetical context: potential recipients of data, potential secondary users of data, but also others that may have in interest in data 'down the line' that are not immediately obvious now.

Furthermore, to what extent should the starting point for assessing (future) identification risk be knowledge based on recorded facts, or can non-recorded personal knowledge also be presumed to lead to identification (at least abstractly)? To what extent should such knowledge be formed upon a plausible basis (indicating a degree of reasonable certainty in its veracity) to present a significant identification risk, and against whose standard of plausibility and knowledge basis?

3.3.1.3 The theory of identifiability underpinning existing concepts of personal data can appear doctrinally useless and outdated

While the problems outlined are less relevant to the notion of 'identified', still clouding that term's interpretation is the fact that the verb 'to identify' (as well as the noun 'identification') can connote a unique singling-out in a group. As mentioned, the implications of such a reading is that the boundaries of personal data, the processing of which should be protected in law, becomes (or may be perceived as becoming) virtually limitless.

For this reason, Korff argues that the DPD's personal data definition suggests, "*any data that can conceivably be linked to an individual (in whatever way, by whoever) [can] be regarded as personal*".[221] Hon et al agree: ultimately, any/all datum which is linkable to a (living) individual is potentially personal data under the DPD, "*because it can identify them if combined with enough other information*".[222] Bergkamp/Dhont also think this lack of clarity in interpreting the identificatory requirement problematic.[223]

While many now agree that the theory underpinning the identificatory requirement is a risk-based one,[224] disagreement remains over how, and to what extent, to assess the degree of identification risk required for data relating to an individual to be deemed personal data. This uncertainty closely links to the tendency to construe the Means Test as relativistic. For Zwenne, for example, the degree of risk involved is (more of) a moot point because *"[t]he aspect of relativity stands in relation to the aspect of reasonableness*" and *"[a] reasonable amount of effort for one does not have to be the same for another*".[225]

---

[221] Korff (2013, p.114).
[222] Hon et al (2011, p.217).
[223] See, e.g. in their analysis of the terms 'identified' and 'identifiable', the argument by Bergkamp & Dhont (2000, p.74) that their meaning is by no means clear. See also Wong (2013, p.39): "*[t]he broad definition of personal data has raised more questions about its specific scope. It is arguable, in my view, that there is a lack of guidance from the Data Protection Authorities concerning the precise nature of the notions "identified" and "identifiable"".*
[224] As Tene points out (2011, p.7), adopting an approach that restricts the scope of the term personal data based on the risk of identification *"confirms to the spirit of Recital 26 of the Directive"*. Notable scholar-advocates of a risk-based approach are: Hon et al (2011); Ohm (2010); Schwartz and Solove (2011), (2013, #1), (2013, #2); and Tene (2011).
[225] Zwenne (2013, p.4). In other words, it requires consideration of what means are likely to be used by someone in practice, but also what means are reasonably likely to be used by them relative to their situation, which will all depend on the circumstances.

Yet, simultaneously, the Means Test has been interpreted as requiring - at least partially - an assessment of objective factors (e.g. the emerging status/availability of identification-enabling technologies generally - i.e. that could be used by anyone, so motivated, assuming feasibility/cost-effectiveness).[226] This partial objectification of the Means Test suggests that the degree of identification risk required will often be lower than one might expect under a purely relativistic test, even without including certain presumptions about deeming data to be personal data for caution's sake. This is borne out by the some of the interpretive evidence cited above that, where there are means reasonably likely to be used to effect identification of a data subject, the likeliness of identification actually taking place can be low in practice and yet data still deemed personal.[227]

In practice, this inter-dependency of factors ('identifiability how', 'from whose perspective', and 'with what likelihood') suggests that determining whether identifiability from data exists under the DPD does not add very much as a limiting restriction on the scope of data protection law. Not least, it is easy to fall back into a 'just-in-case' highly risk-averse model of future possibilities in practice.[228] Moreover, these problems/uncertainties are reflected in national data protection laws that falter in providing consistent guidance on how to assess identified/identifiable. To illustrate, out of those countries that believe determining the identificatory requirement involve assessment of contextual variants in practice, no precise consensus is found regarding which variables should be assessed. Moreover, tackling the complexity of sub-definitional requirements is often side-stepped by national courts. Greater clarity is needed. In the next section, the changes being introduced by the GDPR are considered, along with the extent to which they address such concerns.

3.3.2 To what extent does the GDPR address these concerns?

To determine the adequacy of the GDPR's identificatory-approach, we need to pin-point how it changes the DPD's identificatory requirement, and whether such incoming changes better address the above-discussed challenges related to achievement of the twin aims.

---

[226] As mentioned and for further discussion in Chapter 5, the 'motivated intruder' test works by postulating the identification capabilities of a third party (a hypothetical adversary bent on identifying the data subject from the relevant data), including in respect of capabilities that may not be owned by the data controller or a data recipient in reality. Construed in this way, the Means Test (and its reasonableness factor) can be interpreted more objectively.

[227] Albeit that the WP has also stated (WP136, p.18) that the risk of identification is often so low that it may be justifiable to apply data protection rules more flexibly than if information on directly identifiable individuals (e.g., individuals identified by their name) was being processed: *"[r]etraceably pseudonymised data may be considered as information on individuals which are indirectly identifiable. Indeed, using a pseudonym means that it is possible to backtrack to the individual, so that the individual's identity can be discovered, but then only under predefined circumstances. In that case, although data protection rules apply, the risks at stake for the individuals with regard to the processing of such indirectly identifiable information will most often be low, so that the application of these rules will justifiably be more flexible than if information on directly identifiable individuals were processed".*

[228] The rationale for this argument is explored further below, e.g. with regards to the likelihood of profiling activities being applied to certain data relating to persons in the future increasing the likelihood that its data subjects will later be identified or at least become identifiable from such data.

As a starting point, in 2010 the European Commission noted in a Communication*:*

> [A] consequence of such a broad and flexible approach is that there are numerous cases where it is not always clear, when implementing the [DPD], which approach to take, whether individuals enjoy data protection rights and whether data controllers should comply with the obligations imposed by the [DPD].[229]

Partly for this reason a regulation rather a directive was chosen as the legislative instrument to introduce EU law data protection reform (per Chapter 2) to remove some of the legal uncertainty and fragmentation exacerbated by the DPD (through the risk of divergent approaches that nation-by-nation legal implementations of a directive causes).[230] The GDPR's direct applicability introduces one single legislative text – and one EU-wide concept of personal data replacing domestic legislative concepts of the same (including under the DPA) - should reduce national legal fragmentation and provide greater legal certainty by introducing a harmonised set of core rules. Notwithstanding, its revamped concept of personal data should still aim to ensure effective protection of individuals' rights that might arise when information relating to them is processed, and minimise disagreement over its interpretation (potentially resulting from the fact that linguistic interpretations of the key concepts set out in the GDPR's identificatory requirement could still diverge between MSs).

The GDPR's definition of personal data bears strong resemblance to the DPD's definition, but with several textual improvements.[231] It adheres to the same four-pronged requirements approach. Article 4(2) states:

> 'personal data' means any information relating to an identified or identifiable natural person 'data subject'; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, **location data, online identifier** or to one or more factors specific to the physical, physiological, **genetic**, mental, economic, cultural or social identity of that person" (emphasis added to indicate material changes to Article 2(a) DPD)

The GDPR appears to broaden the coverage of data protection law by providing additional data-identifier examples – location data and online identifiers – and one extra ("genetic") factor by which an individual may be identified/identifiable. The DPD's encompassing of these new additions may

[229] European Commission (2010 Communication, p.5).
[230] In other words, the GDPR's introduction removes the need for implementing domestic data protection legislation reducing national differences to a minimum to avoid eroding the principle of harmonisation.
[231] The use of the word "improvements" here seems appropriate taking into account that (per Chapter 2) the twin objectives of the DPD are also adopted by the GDPR, albeit that – when it comes to the definition of personal data under the GDPR – there is no substantial 'reinvention of the wheel' from the model adopted under the DPD.

already exist implicitly. However, explicit confirmation under the GDPR that identification is considered achievable (legally) by linking a particular person to online identifiers, and location data, associated with devices used, is useful. In that context, Recital 30 GDPR also encourages data controllers to carry out assessments to consider whether steps to combine such identifiers with other information are likely to be taken:

> Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. **This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.** [232] (emphasis added)

Together, Article 4(2)/Recital 30 GDPR suggest that online identifier-data associated with devices can be deemed personal data (through association with particular device users). Yet there should be no automatic assumption of data-identifiability from online identifiers, albeit profiling activity is one obvious way in which personal identification may be achieved. Emphasis on context-dependent interpretation appears at the heart of the identification risk-assessment process for determining when personal data is present under the GDPR.

Assessed against data protection's aims, this position could lead to an unduly restrictive interpretation of personal data (denoting less privacy protection) when compared to a counterfactual assumption that – say - all IP addresses are personal data. Alternatively, it could also be construed as leading to an expansive interpretation of the notion of personal data (denoting more privacy protection) where IP addresses are determined to be personal data when otherwise they might not be. In other words, the identificatory-approach can support a very expansive interpretation of the boundaries of data protection law (in the main text) or a very narrow one (in the recitals). Which interpretation is correct depends on identifying the appropriate counterfactual perspective controllers are meant to adopt (currently an issue of debate). The gap between the two possibilities (very wide or very narrow) conveys the wide extent of potential discrepancies between decision-making in this area on the facts. Either way, the benefits of retaining a context-dependent analytical approach are apparent: what is lost in legal certainty on application of its approach

[232] Per Chapter 2, the risks associated with online activity have been one of the major incentives to suggest the revision of the existing regulatory framework of data protection. This is shown by the inclusion of the reference to "online identifiers" in Recital 30 and Article 4(1), as well as the reference to online tracking/profiling in Recital 24. Thus, the GDPR seems to recognise that data collection and processing through online tracking techniques pose distinctive privacy threats and should be covered by data protection law (Article 3(2)(b)).

according to the facts at hand (assessed against many open-ended factors) translates into certain gains because that test is inherently flexible.[233]

Recital 26 GDPR retains a version of the Means Test. It states:

> The principles of data protection should apply to any information concerning an identified or identifiable natural person….To determine whether a 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 any other person to identify the individual directly or indirectly. **To ascertain whether means are reasonable likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development**...[234] (emphasis added to indicate material changes to Recital 26 DPD)

Clearly, continuity is intended between Recitals 26 DPD/GDPR regarding guidance on determining identifiability. The Means Test remains virtually identical as a non-binding standard. Consequently, this provides greater support (and more certainty) for assessing identifiability broadly in line with previous EU guidance on this point.

Notable differences include the explicit expansion under the GDPR of the 'identifiable' notion to include singling-out as a possible identification means.[235] This insertion - suggesting that the GDPR's rules (may) apply to data processed from which particular individuals may be distinguishable from others - confirms that a high level of protection for individuals' rights is intended. However, Recital 26 GDPR falls short of stating that identification automatically encompasses individualisation (thus unique identifiers should always be deemed personal data) exhorting only that singling-out be taken into account as a means reasonably likely to be used for identification.

Therefore, questions remain on the degree of acceptance in the future under EU data protection law (and policymakers' interpretations thereof) that a person be considered identifiable when, within a group, they can be singled-out without their 'real' identity being known, including via

---

[233] Schwartz & Solove (2013, #2, p.13): *"[m]uch depends on judgments about open-ended factors, such as "the means likely reasonably to be used".*

[234] Compare, in the DPD, the exhortation that 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. The remainder of Recital 26 GDPR, as well as references to pseudonymisation in the GDPR, are discussed in Chapter 5 below.

[235] As discussed earlier in this chapter, a person may be considered identifiable when, within a group of persons referred to in a dataset, they can be distinguished from others (without their 'real-world' identity necessarily being known).

online identifiers.[236] This confusion appears in the WP's last letter to the EU Institutions during the GDPR trilogue negotiations, where it annexed the concept of being able to single-out an individual from data to the fact that they could therefore be treated differently subsequently.[237]

Regarding the Means Test, Recital 26 GDPR exchanges the phrase "means likely reasonably" (from Recital 26 DPD) with "means reasonably likely". Also added are explicit reference to the taking into account of "*all objective factors*" including "*the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development*". Therefore, the consideration of subjective factors only is rejected.

Despite such changes, problems with interpreting the Means Test persist under the GDPR. When we drill down into the phrase "means reasonably likely", uncertainties remain:

- First, how the GDPR Means Test should be interpreted upon practical application remains unclear. It exhorts the consideration of all those means for identification likely reasonably to occur, but what then? How should we read the 'taking into account' part of the test? Does it prescribe assessing each individual means individually - or perhaps "all" such means as a whole cumulatively assessed somehow – i.e. how should an account taken translate into an ultimate decision that data are personal?

- Second, the relevance of subjective factors; to what extent they should be considered and which factors exactly (such as processing intent/purpose) remains unclear.[238] Furthermore, emphasis upon "*all objective factors*" as well as the trend for endorsing a hypothetical (motivated intruder) perspective on identifiability – objectifying the party that could achieve identification and their capabilities - suggests a relativistic approach to personal data (viewed from one perspective) may not be the correct one. Yet, applying the GDPR

[236] For further discussion in Chapter 5, the GDPR also refers to the fact that data which has been through a pseudonymisation process can still be used to single-out individuals via unique indirect identifiers left in the data.
[237] See, e.g. Article 29 Working Party. Letter to Ms Ilze JUHANSONE Ambassador Extraordinary and Plenipotentiary Permanent Representative to the EU, Brussels, 17 June 2015, p.2: *"[t]o ensure the general objective of maintaining a high-level of protection of personal data is upheld, personal data should be defined in a broad manner in line with technological evolution. The definition of personal data should therefore take into account the situation in which people can be "singled out" on the basis of identifiers or other information and could subsequently be treated differently*". See also comments by the WP on the GDPR in draft form (Article 29 Working Party (2012, WP191) where it advocated a more explicit introduction of the concept of pseudonymisation in its main text, such as its inclusion into a broadened definition of personal data (p.11). As such, the WP appears to have been endorsing more recognition of an overt kind in the law of the fact that there is a spectrum of risk, from the acceptable to unacceptable, and that those enacting data protection compliance programmes should consider such factors when setting their data strategies. At the same time, the WP recommended introducing a general obligation to anonymise or pseudonymise personal data ("*where feasible and proportionate according to the processing*") into the draft GDPR's Articles 5 ('Data protection principles') and 23 ('Privacy by design') (p.11).
[238] Although, to note, the WP has stated (in Article 29 Working Party, 207, p.17): "*[w]hile it is useful to think through the potential motivations of …potential intruders, the WP29 emphasises that there are also considerable limits to this approach: The exercise may be to some degree speculative. In the absence of obvious 'motivating factors' such as those described above the exercise may lead to false reassurances…"*

Means Test practically still appears to require considering relevant facts as they are. There seems to be fundamental tension between these instructions.

- The Means Test remains spurious in practice when applied to unknown third parties, and unknown additional information 'out there somewhere', and future scenarios (around data creation, availability, and access), at least while uncertainty persists regarding the correct perspective on identifiability.

Hence, while the GDPR addresses some existing problems with the identificatory-approach, challenges remain. Its broad language retains the inherent flexibility of the identificatory requirement in theory, but practical uncertainty persists. Specifically, it is unclear what the GDPR directs data holders to do exactly regarding the non-binding Recital 26 Means Test. Moreover, the GDPR's identificatory concepts – rather than clarifying current interpretive confusions – compound them by leaving mostly unaddressed interpretive-polarities on identifiable how, from whose perspective, and with what likelihood.[239] Questions remain as to precisely what factors need to be assessed for determining when data can be said to be identifiable of someone, using means reasonably likely to be used, and how all such means identified should exactly be taken into account to reach a final conclusion on whether data are personal or not, including regarding the degree of identification risk required.[240]

3.4 Chapter conclusion

While Chapter 2 argued that the intention of including an identificatory requirement in legislation was to delimit the outer contours of the personal data concept,[241] this chapter demonstrates that

[239] An indicator of this uncertainty can be seen in the varying approaches to the definition of personal data put forward by the EU Institutions in draft versions of the GDPR during reform discussions, where proposals for change in this respect were both structurally and substantively significant. For example, in the European Commission's 2012 text version of the GDPR it proposed raising the importance of the Means Test by placing it in the main text of the regulation. The Commission proposed that personal data should be defined in the GDPR as: "*any information relating to the data subject*" and the concept of data subject defined separately as an identified natural person or a natural person (draft Article 4(2)): *"who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person".*

[240] This ambiguity manifests itself in different legislative interpretations of the identificatory element in different EU MSs' laws that can make for differences in the operation of national data protection law regimes. The consequences of this ambiguity are explored further in Chapter 5.

[241] To recap, without the limiting functionality of the identificatory criterion attached to the concept of personal data under EU data protection law, it has been argued that postulating personal data only requires that a piece of data for processing is about a specific (living) person. That is, under this counterfactual scenario, the application of EU data protection law would only require the supposition of some sort of loose nexus between information that is to be processed and a specific person. This statement assumes that the 'relating to' criterion under EU data protection law is meant to be interpreted very broadly - indeed as a low-level, catchall standard of informational association - as explained in Chapter 2.

legal notions of identified/identifiable (under the DPD, DPA, and GDPR) are ambiguous in critical respects. While creative attempts to mould their linguistic 'open texture' in response to technological developments, this has led to divergent interpretations. Not least, there is underlying disagreement over what the grounding verb 'to identify' actually means in data protection law (with alternative positions diminishing[242] or elevating the protection of data subjects' rights where data relating to them are processed) and over the extent to which it should be contextually/dynamically assessed (as opposed to reliant on a priori attributes).

Incongruous EU interpretations are liable to impede information flows, resulting in practical contradictions (especially in grey areas, e.g. IP addresses) in the past, present, and future, in turn potentially undermining the level of rights' protection for individuals when data relating to them are processed.[243] The binary nature of the personal data concept exacerbates this uncertainty. Yet, the often-graduated transition of some data from non-identifiable to identifiable (and vice-versa) makes cut-clear distinctions about personal data status very hard.

Bringing into question the workability of the identificatory-approach in the next decade with the development of the IoT, and beyond, is also the rapid growth of corollary information relating to us 'out there somewhere'. In addition, the prevalence of secondary uses of data relating to persons will become standard, as well as the unknown nature of additional information potentially in the possession of a third party accessible by the data controller at some point. Thus, once-and-for-all time decisions about whether data protection law applies are unrealistic, but periodic assessments of identification risk are no less challenging.[244]

Chapter 4 considers sub-research question 2, defining the contours of an alternative approach and its strengths/weaknesses assessed against the twin objectives.

---

[242] If it is interpreted narrowly, the term personal data could be restricted to data that is capable of identifying an individual in real-life, either by itself or in combination with other data. Identification, in this context, could be direct or indirect. However, per Tene (2011, p.6): "*[a] narrow definition of personal data may fail to account for the increasingly sophisticated means of identifying"*, including via singling individuals out from others in a group to whom the data relates. Compare, Dinant (2010, p.22, translated into English): "*[w]hat is important now in the new technological context is rather individualisation than identification. …Without even asking about the "identity" of the individual, that is to say their name and address, we can characterize them in terms of socioeconomic, psychological, philosophical or other criteria and enforce certain decisions on them to the extent that the individual's contact point (their computer) does not necessarily require the revelation of their identity in a narrow sense. In other words, the ability to act vis-à-vis an individual no longer necessarily requires the ability to know his identity. Such identification may be regularly realised within electronic environments. An example may be provided by those individuals who 'date' online in chat rooms while concealing details of their 'real' identities or physical locations".*

[243] This is in addition to the fact that as stated by Booth et al (2004, p.123): "*indirect identification no longer appears resident within the nature of the data itself, but rather within the nature of the context within which the data resides … Whether a specific piece of data will enable identification either directly, or indirectly, thus depends upon whether sufficient information is already present to attribute the identifier with 'unique status' and to 'directly' enable identification".* In other words, it is being argued that the incongruity is caused mainly by the lack of a coherent and clear theoretical framework governing how context-dependent assessments should be carried out.

[244] The crux of such challenges relates to data accretion problem, which Schwartz & Solove (2013, #1, pp.913-4) pinpoint as *"the moment at which information becomes identifiable enough to fall within the scope of a particular law".*

Chapter 4 - The Effects-based Approach to the legal concept of personal data

The identificatory requirement and its key functionality in contouring and fixing the outer boundaries of the personal data concept (per Chapter 2) could be replaced with an effects-centric criterion. Briefly, this would require a 'Relevant Effect' (described below) likely flowing from the processing of 'information' 'relating to' 'a natural person' as a pre-condition for data to be deemed personal under EU data protection law.[245]

Extrapolating on the possible content, benefits, and parameters, of this approach in a particular model form as proposed - hereafter termed 'the Effects-based Approach', which could form part of a new legislative definition of personal data - forms the subject matter of this chapter. It addresses sub-research question 2 regarding the effectiveness of an effects-based approach to personal data in terms of realising and reconciling the twin goals of facilitating the free flow of personal data, and safeguarding individual rights:

> **Under an effects-based approach to personal data (exemplified by the statement, 'information can only be personal data if it is processed in a manner capable of affecting the individual to whom it relates appreciably'), how effective could this approach be in terms of realising and reconciling the twin goals of facilitating the free flow of personal data, and safeguarding individual rights?**

In terms of chapter structure:

- **Section 4.1** describes the features and rationale for the Effects-based Approach in model form, together with the relationship between the Effects-based Approach and risk-related terminology.
- **Section 4.2** analyses doctrinal support for the Effects-based Approach in existing EU guidance on the concept of personal data.[246] Examples include some important sources of

[245] In other words, it proposes to leave the remaining three legal element-requirements found with the EU data protection definition of personal data intact: "any information"; "relating to"; and, "a natural person".
[246] Per Chapter 2, this analysis is deemed worthwhile - in addressing the research question (and sub-questions) - on the basis that a high degree of doctrinal support for the Effects-based Approach evidenced within existing data protection law may be considered more likely to facilitate an increased level of pan-EU legal certainty in it if it were introduced. Not least, a high degree of consistency promotes legal and practical certainty for economic operators and individuals, as well as data protection agencies and other public authorities.

legislative interpretation suggesting effects-based foci in the DPD, the DPA, as well as the GDPR.

- **Sections 4.3 and 4.4**, respectively, assess the Effects-based Approach as modelled and its compatibility with, and its potential to realise and reconcile, the twin aims of data protection law. In this respect, a critical assessment of limitational factors that could constrain the Effects-based Approach – specifically, to shape the notion of 'Relevant Effect' – is set out. These sections also highlight how the Effects-based Approach might address some of the shortcomings of the existing identificatory-approach to personal data under data protection law.

- **Section 4.5** concludes the chapter.

4.1 Explaining the Effects-based Approach: what it is, and what it is not

4.1.1 Risk and risk-assessment

The effect-based approach shares an association with notions of risk and risk-assessment, requiring brief explanation. While similarities are noticeable straightaway here compared with so-called 'dynamic' interpretations of the identificatory requirement per the last Chapter (see pp. 74, 78, and 80 above), it is useful to go back to basics regarding risk terminology and meaning summed up as follows:

- Risk is a broad concept associated with the likelihood of something happening in the future connected with a given event or action/inaction.

- Typically, this possible occurrence ('the risk that something might happen', whatever that may be) is one with negative implications ('the risk that something bad might happen').[247]

- The notion of risk-assessment, by comparison, normally implies an ex-ante (upfront) analysis of what might happen in terms of a spectrum of likelihood (the weighing of the likelihood that something bad might happen, which can be contrasted against an assessment of the likelihood that it might not). Said otherwise, there are two kinds of uncertainty in risk management. One is uncertainty of outcome, and the other is uncertainty of probabilities.

[247] Thus, the Oxford English Dictionaries [online] defines risk as, "*A situation involving exposure to danger*". Compare the Merriam-Webster Dictionary [online] definition of risk as, "*the possibility that something bad or unpleasant (such as an injury or a loss) will happen*".

- While some of the likelihood ('that something bad might happen') may be foreseeable and/or foreseen, not all may be. Risk is often associated with the consequences of action taken in spite of uncertainty of outcome.[248]

Another key point about the concept of risk (events-predictive) is that it is often confused with 'impact' (harm-consequential to events-predictive and, specifically, the likely magnitude or severity of that harm). For example, the OECD says, *"'[r]isk' is intended to be a broad concept, taking into account a wide range of possible harms to individuals*".[249] The two points are distinguishable in this thesis: a risk of something negative occurring could refer to the possibility of a particular impact consequence (i.e. the risk of harm occurring). However, while the former is a condition precedent to the latter, it is not necessarily consequential.[250] Notwithstanding, risk-assessments typically involve assessing the likelihood of an adverse outcome occurring, as well as assessing the likely severity or magnitude of that adverse outcome if it happens. Thus, to understand fully the nature of a risk in a particular context, consideration of both risk type ('a risk of what exactly?') and the possible impact of that risk ('a risk of what exactly occurring with what negative effects exactly?') seems inevitable.

4.1.2 Inspiration

The catalyst for developing the Effects-based Approach by this thesis author came from considering the rationale for expanding the identificatory requirement to encompass singling-out individuals from data, specifically in recognition of singling-out possibilities - and the possible underlying interests deemed worthy of protection – in digital scenarios.[251]

It also encompasses consideration of substantive concerns underlying the prevailing debate mentioned in the last chapter about whether – and, if so, when – IP addresses (particularly dynamic IP addresses) are personal data. If not treated as personal data, for example, this could pose problems in light of the risks of something bad occurring that the subsequent unfair use of such data could represent for the subject-individual (e.g. linked to profiling activities).

---

[248] Thus, one definition of risk is "*the intentional interaction with uncertainty, which is a potential, unpredictable, immeasurable and uncontrollable outcome*" (Dinda, 2016, at p.363).

[249] OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980, rev.2013, p.24).

[250] In particular, the risk of identification of a person from data in the future refers to a possible, factual consequence – viz. a technical fact - not an impact one, albeit that the latter may seem implied from the former.

[251] In other words, what exactly is the concern that presages the involvement of data protection law in the processing of data where this consequence of singling-out will or could occur (but the revelation of the 'real-world' identity of the individual is unlikely)? To note, this touches on one of the 'grey identifiability areas' noted as concerning in the last chapter, such as in respect of online identifiers, in relation to which scholars such as Schwartz & Solove (2011, p.1814) have mapped out new identification-based models in reconceptualising the US concept of PII.

Several candidates for such concern (the risk of 'what bad thing exactly' might happen because of the singling-out of an individual from data) present themselves, i.e. as candidates for the logical and conceptual basis upon which to reshape the test for personal data. For example, because of singling-out from data, an individual may become contactable by others.[252] Accordingly, one could postulate that the outer edges of the concept of personal data should encompass situations where processing information relating to individuals would (as a technical possibility) enable them to be (at the least) contactable by third parties.[253] Notwithstanding, a criterion for personal data to exist requiring the linking of a one-to-one communications bridge with an individual appears too narrow. Intuitively, whether fair processing protection is required (in respect of processing data from which individuals can be singled-out) seems tangential to the issue of whether someone can contact them or not from such data.

In a similarly broad context, potentially concerning is the ability for digital marketers to subject unknown individuals to advertising exposure consequential to singling them out from data. The underlying worry here seems to be that an individual may be targeted and treated differently from others.[254] However, in begging immediately another question - 'treated differently in what ways exactly?' - it appears that postulating a new criterion of 'resulting-in-a-difference-of-treatment-of-the-data-subject-compared-to-other-individuals' enabled through data processing (in lieu of the identificatory requirement) would also fail to provide an effective functional limitation upon the outer parameters of the concept of personal data.

In discarding those approaches as falling short of reaching the essence of the concerns raised by the prospect of singling-out 'unknown' individuals, the Effects-based Approach construes this problem as relating to the prospect of negative effects (impact) upon the individual to whom the data relates caused by a particular processing activity applied to that data. By extension of this argument, it is this prospect for harm to an individual that also appears to justify, at base level in theoretical terms, why the attribute of their personal identifiability from data merits a data status worthy of the law's protection.

---

[252] In illustration, in the digital domain, a person may still be distinguished uniquely in circumstances not revealing of their civil identity where their unique email address is processed. In turn, knowing this email address in theory lets a third party communicate with the holder of the email account.

[253] See arguments in Gratton (2014, p.177).

[254] Compare, Article 29 Working Party. Letter to Ms Ilze JUHANSONE Ambassador Extraordinary and Plenipotentiary Permanent Representative to the EU, Brussels, 17 June 2015, p.5: *"[t]o ensure the general objective of maintaining a high-level of protection of personal data is upheld, personal data should be defined in a broad manner in line with technological evolution. The definition of personal data should therefore take into account the situation in which people can be "singled out" on the basis of identifiers or other information* ***and could consequently be treated differently"*** (emphasis added). See also, ibid, Appendix, p.5: ***"****[a] natural person can be considered identifiable when, within a group of persons, they can be distinguished from others* ***and consequently be treated differently****. This means that the notion of identifiability should include singling out individuals"* (emphasis added).

Under the Effects-based Approach, therefore, rather than focusing on the nature of data in a static sense, this definitional lens would emphasise the potential consequences of data's usage in a dynamic sense. It would affirm the connection between the prospective impact flowing from a processing activity applied to data relating to an individual upon them, and whether the processing of such data should therefore be deemed sufficiently concerning to be subject to data protection rules or not in that specific processing activity context. Said otherwise, data that relates to a person would be considered personal data when it is assessed as likely to have a 'Relevant Effect' on that individual following a particular processing activity.

Thus, further development of this concept of 'Relevant Effect' – what is required by way of negative impact flowing from data to become personal data legally in relation to specific individuals - is necessary (see Section 4.4 below in terms of considering minimum theoretical contours of its individual components). For example, further specifications are required in terms of the type and magnitude of effects considered sufficient for a Relevant Effect to be present, as well as the minimum degree of likelihood that such an effect might arise.

For now, suffice to say that evaluation of the data context is a crucial aspect of limiting the Relevant Effect sub-concept.[255] Specifically, its finding would depend on whether the negative effect liable to flow from the processing of data falls in or out of a continuum of effect upon the individual to whom it relates in the particular instance under consideration. Therefore, any risk-assessment (prediction) as to the likelihood of a negative effect occurring must take into account the particular circumstances of the data processing under consideration. Furthermore, all data would be assessed against the background of the usage context.

4.1.3 Choice of thesis terminology

As mentioned, a negative/adverse impact upon a person is being equated to that of negative/adverse effect, both of which are also construed interchangeably with the concept of consequential-harm.[256] Accordingly, the reference in the Effects-based Approach to the concept of

[255] Like the Means Test standard, this reflects a belief that personal data is not categorically deserving of protection, but instead that appropriate protection is dependent upon the data context. That is, both models adopt a contextualised ('context-oriented') definitional view of personal data meaning that data becomes personal according to its context.
[256] Thus, the noun 'effect' connotes the requirement for a consequence, which in relation to a particular person implies that in order for someone to be 'affected' by something, it must have some form of impact upon them. Whereas 'harm' is defined in the Merriam-Webster Dictionary [online] as, "*physical or mental damage or injury : something that causes someone or something to be hurt, broken, made less valuable or successful, etc."* In other words, it suggests requirement of some tangible negative effect, in the sense that it results in damage that may be suffered by someone palpably (either physically or mentally). Of course, effects/impact/harm can be referenced as being suffered in actuality, or referenced in respect of a likelihood that they may be suffered by someone.

personal data might also have been termed either an 'impact-based' approach, or a 'harm-based' approach.

The reasons for not choosing these alternative labels are as follows:

- The term 'harm-based approach' has been associated with the definition of personal data, but in a narrow sense referring only to ex-post determinations of damage caused by data processing – that is, solely focused on negative effects that have actually been caused and evaluated as such *after* the processing takes place.[257]

- The term 'harm-based approach' is also found in the context of describing a school of thought contending that the focus of privacy and data protection regulatory activity should centre broadly on data usage and its likely consequences.[258] However, these and similar scholarly proposals are not definitional, which may cause confusion with the definitional (i.e. jurisdictional) hypothesis put forward here.

- The term "impact-based approach" risks confusion with later references in this chapter, and later chapters, about privacy/data protection impact assessments (albeit a connection is made between these concepts and the Effects-based Approach).

4.1.4 Interdisciplinary literature overview and how non-legal scholarly theories situate alongside a legal proposal for an effects-based approach

As a Web Science influenced piece of research, it is appropriate to include a section providing critical engagement with a selection of cross-disciplinary literature dealing with issues discussed in this thesis, in particular to explore some resonant theories to an effects-based approach from non-law disciplines.[259] This overview will help to disentangle the likely consequences (intended or unintended) of the main thesis ideas put forward, as well as help the reader distinguish the main strands of the new theory set out in this chapter and give some indication of how it will be developed in later chapters.

[257] For further discussion in Section 4.3 below and Chapter 5.
[258] See ibid.
[259] As mentioned in Chapter 1, discussion is kept brief on concepts of privacy from a non-legal perspective (see pp.143-144), although interdisciplinary analysis related to the concept of identification is contained in Appendix 1.

Bamberger/Mulligan[260]

Bamberger/Mulligan identify drivers for privacy-friendly corporate behaviour through observing how companies manage privacy 'on the ground' and the intensity of privacy management practices employed. Their review leads them to conclude that a sympathetic corporate culture is the most important factor in motivating organisations to embed privacy into corporate management and business operations.

This perspective is a reminder that any analysis of the effectiveness of local laws in this area – whether highly regulated or not – is best underpinned by empirical inquiry into the behaviour of those tasked with protecting privacy within organisations. Said otherwise, a corporate culture of privacy management is shaped, not just by the regulatory environment, but also by factors such as the recent rise of privacy-bespoke professionals in organisations (prompted, in turn, by increased media pressure for more privacy protection with socio-technical advances).

The authors draw best practices and policy recommendations from their organisational approach to privacy governance. These are focused on efforts to increase opportunities for 'bottom-up' self-regulation and institutional engagement. To some extent, the introduction of the GDPR (with its requirement to install data protection officers, data protection by design/default, and the deterrence effects of its higher fining levels) introduces change to the same effect. As acknowledged elsewhere in this thesis, the full extent of the GDPR's influence on the regulatory environment and corporate compliance remains to be seen in the next few years. Notwithstanding, this thesis recognises that there is still need to develop new regulatory incentives to change behaviours within organisations, including via the introduction of compliance tools adaptable in anticipation of future privacy-challenging problems beyond the GDPR.

In the above sense, changing laws is just the start. The hard work comes in devising and making accessible ('on the ground') mechanisms to help people exercise meaningful control over the use of data relating to them, and to help organisations simplify their compliance efforts in ways that promote regulatory coherence and uniformity. In other words, policy makers cannot rest on their laurels with the GDPR, but will "*need to think harder in the years to follow and come up with better and bolder solutions that are capable of reconciling individual fundamental rights and the increasing demand for free cross-border data flows* ".[261] It is in that context, the proposal for a new regulatory

[260] Bamberger & Mulligan (2015), who carried out a five-nation study of corporate privacy management in the US and Europe (including the UK) identifying international best practices and making policy recommendations.
[261] Burri & Schär (2016), p.507.

effects-based (block exemption/safe harbour) tool introduced in the next chapter should be understood.

Abelson et al[262]

Abelson et al advocate usage-oriented privacy regulation in the context of discussing interdisciplinary solutions that could be employed to augment Web architecture in view of societal needs. Said more simply, they suggest that policies to support the development of the Web should be based upon an increase in accountability and transparency around how information is used. Conversely, they argue that when information is used inappropriately, it should be possible to pinpoint precisely what happened.

The authors propose that more accountable data usage would require, not only augmenting the metadata of information used online with usage policies, but also creating technical mechanisms in support. The latter, they say, could provide machine-readable representations of policy rules to facilitate compliance when information is used. However, it is also acknowledged that this proposal raises practical challenges regarding the deployment of such information-accountable technologies on the Web and in enterprise systems, such as through protocol changes. The authors suggest using royalty-free technical standards (e.g. a technical language provided by the non-regulatory body W3C to allow data creators to explain preferences about third party data usage in metadata) so they can be implemented widely with ease.

The above perspective highlights that substantial technical challenges loom to design systems that constrain the misuse of information (e.g. around auditing and enforcement). Another big challenge is the limitations of rule-reasoning, viz. how to create rule-restrictions on the use of information, and how restrictions can interact when different pieces of information are combined. To this end, the authors argue that 'policy-aware systems' should be able – inter alia – to mediate access to data sources and maintain logs of data transfers, *"presenting accountability reasoning in human readable ways, and allow annotation, editing, and publishing of the data and reasoning presented*".[263]

Like Abelson et al, this thesis also envisages modelling an accountability-focused compliance tool that can sit easily besides policy and systems architecture frameworks under development. It is clear that laws and systems design must accommodate each other, particularly where both are oriented information accountability (e.g. to deter personal information being used in ways that have negative effects), and also help incentivise accountable data management. Thus, legal and

[262] Abelson et al (2008).
[263] Ibid, p.7.

technical modelling can support each other but each needs to understand how the other works. Awareness of and support for new strategies for privacy protection are also important for information accountability to work in practice.

While the effects-based model associated with the (block exemption/safe harbour) tool ultimately put forward in this thesis does not immediately providing answers to the underlying questions associated with identification-risk in a big data era, it should be situated within a comprehensive overview of legal/technological/organisational insights adaptable to new technologies and evolving societal norms. Such insights present a direction of travel towards greater accountability through incremental changes, via multiple inroads that will need to be deployed and must be capable of working successfully in overlap.

Hartzog[264]

Hartzog believes that the existence of many different kinds of privacy harms (in terms of clarity, directness and severity) should embolden lawmakers to acknowledge the need for equally diverse solutions for protecting personal information. He argues that these protections do not always need to be substantial or direct, such as by wholesale changes in the law (which are not always suitable), but can be relatively modest privacy protections that are very effective incrementally in providing adequate and balanced remedies.

The current value of such modest protections is underappreciated particularly, says Hartzog, regarding the potential for stemming harms resulting from the misuses of semi-public information of an opaque, novel, remote, and cumulative nature. He says these protections provide a source of privacy regulation that can be conceived dynamically (to fill in gaps where law does not or cannot provide recourse) at an overarching level, and which is flexible enough to focus on context-specific concerns. For Hartzog, focusing on incremental, modest safeguards is also useful in trying to balance privacy protections with other values. Specifically, it forces lawmakers to prioritise the harms requiring the protection of the law, while also inducing them to provide clearer rationale for why a law was introduced, and the remit of its application (e.g. in situations where legal privacy protection could be reduced because of the presence of other values, such as free speech).

In the same manner, as stated above, the thesis model suggests a piecemeal privacy protection model as a kind of nuanced legal response – as one part of a holistic arch of responses to privacy harm (such as systems design-based protections). It emanates from one particularly challenging area: determining the outer jurisdiction of data protection law, and how to constrain it in a

[264] Hartzog (2014), p.333.

meaningful, but also practical, way that balances it with competing values. Under an effects-based approach of the type put forward in this thesis, such values are equated with the recognition of positive effects that can come from processing data relating to people; whereas, it will be suggested in Chapter 6 that less robust regulation in certain types of processing situations (determined by consideration of the data context, likely harms and likely benefits) would ultimately be the preferred and appropriate approach in the future.

Hartzog/Stutzman[265]

In another article, co-authored with Stutzman, Hartzog classifies obscurity as an example of a type of modest protective measure that can be used for 'good enough' safe disclosure of information not intrinsically confidential or sensitive in nature. Hartzog/Stutzman introduce the concept of obscurity as the idea that information is hard to obtain or understand, although not necessarily inaccessible, because it "*exists in a context missing one or more key factors that are essential to discovery or comprehension*".[266]

More precisely, the authors advocate a legal concept of 'online obscurity', as a helpful guide in privacy disputes. They stress a combination of factors in creating this legal concept, such as search engine visibility, the use of privacy settings and pseudonyms, and limiting the disclosure of information.[267] In that context, the concept of online obscurity is also touted as a practical tool for those processing personal data, by way of a sort of representational 'yardstick' against which it may be determined the extent to which information can safely be disclosed online. For example, Hartzog/Stutzman suggest that it could be referenced in an agreement where online users would be allowed to further disclose information so long as they promise to keep the information generally as obscure as they received it. Alternatively, courts could mandate some form of obscurity on websites instead of forcing them to remove sensitive information.

Online obscurity is a concept that most neatly fits into an identificatory risk narrative. Namely, it aligns with the theory that information – in context - falls upon a spectra of identifiability that runs from a data subject being totally obvious, to he/she being totally blurred/hidden (although still technically possible to re-identify by those with the right abilities and means). However, the concept unpacked also has parallels with an effects-based approach of the type put forward in this thesis in that the continuum described by Hartzog/Stutzman relates to an information 'state' upon disclosure. In other words, it is linked to a data processing activity (involving communications on

[265] Hartzog & Stutzman (2013, #1).
[266] Ibid, p.4.
[267] Information is obscure online, according to the authors, if it lacks at least one of the four key factors that are necessary for discovery or comprehension: search visibility; unprotected access; identification; and clarity.

the social web). Moreover, in that context, the theory focuses on how to minimise the possibility of harm befalling data subjects about whom information obscured relates (in light of the risk that it could subsequently be discovered/comprehended by unintended recipients). Going further, in 2013 Hartzog/Stutzman proposed the concept of 'obfuscation by design' (i.e. obscurity considerations enhancing privacy by design efforts).[268] They suggest that this could be a means to achieve mostly 'good enough' protection from the harms that may result from profiling.

Like Hartzog/Stutzman's concept, the effects-based modelling developed in this thesis is also intended to act as a legal guide to minimising harm in context-specific processing scenarios, while also recognising other values that may flow from potentially beneficial activities. As discussed in Chapter 6, the preferred vehicle – a block exemption - could also specifically be tailored to help those planning to enter into information sharing arrangements. This could provide a more desirable legal outcome than alternative harsher methods (e.g. near total lock-down of any further information usage though the imposition of very harsh agreement terms upon data recipients).

Ambrose[269]

Ambrose challenges the notion that data has a permanent life span and remains static in nature, arguing that the life cycle of data collection and use is often more transient than commonly assumed.[270] Her exploration of this conviction is in the context of identifying specific time-critical points in data processing where enforcing a right to be forgotten online may be considered reasonable (or necessary) in efforts to alleviate harms caused by data accessibility. In other words, she begins to create a taxonomy to help assess competing interests at stake in different time periods in different contexts.

Like the main thesis ideas proposed, Ambrose stresses the fact that data should be conceived dynamically with changing values over time dependent upon the use contexts. According to both, moreover, beneficial outcomes from processing data (e.g. for Ambrose, the benefits arising from online preservation of certain information) is a positive factor in relation to which regulation should be tailored.

Non-legal literature review summary

To summarise, the similarities drawn between the main ideas of this thesis and the ideas in the above literature are indications that a cross-disciplinary perspective is needed on the main thesis

[268] Hartzog & Stutzman (2013, #2).
[269] Ambrose (2012).
[270] Ibid, p.390-1, referring to studies finding that the average URL has a lifespan of 44 days and an average webpage has a lifespan of 100 days.

problems. Contrariwise, a solely legal perspective (that focuses only on changes to data protection/privacy laws without more) would be too restrictive. In particular, the thesis model needs to be compatible with corporate and technical realities, as well as social conventions, now and as they may develop in the future via drivers for change outside legal and regulatory structures. Conversely, in the spirit of Web Science, it is recognised that no one perspective holds all the answers.

Moreover, while legal solutions (such as putting new laws on the statute book) are intended to impart a level of inbuilt certainty associated with formalised rules, their power to guide behaviour through deterrence (i.e. the possibility for sanctions if the rules are not followed) is an unwieldy one. Further factors such as other types of incentives for compliance need to be considered and practicality.

At the same time, as data protection is ongoing a seismic change through modernisation with a focus on increased accountability and transparency, this reform reality must also be encompassed within the main thesis ideas in terms of potential legal responses to the problems outlined in Chapter 1. Yet, any sense that the GDPR solves all the problems is misleading. New and evolving legal responses are needed to support and interpret the main ideas behind the GDPR reform, as we await case-law under the GDPR in coming years, in response to ongoing technological change, and in anticipation of massive increases in the processing of data relating to persons. Likewise, privacy norms and values are transformable according to changing technological and social conditions, with which the law and regulation must keep up in being able to align and balances existing and new interests in data processing.

A note on economic influences on my thesis model

Although legal analysis dominates in the ensuing chapters, economic influences can be found in the Chapter 6 analysis below in the discussion of competition law as a type of modernised regulation founded upon the development of economic theories of harm. It is also suggested that similar kinds of thinking could be useful in helping to develop the regulatory framework presented by the processing of data relating to people and the types of resultant effects (negative and positive) that may be predicted. In particular, the point is made that non-lawyers – specifically economists - may be best placed to analyse and model a taxonomy of data protection harms (on p.257). Notwithstanding, it is acknowledged that for an economist the potential damages from the dissemination of consumer information may be varied and different from those heeded by a lawyer, e.g. to include the decrease in market value of personal data, given its wider availability and lower scarcity.

For completeness, it is acknowledged that contemporary economic ideas have already influenced the informational privacy narrative. A cost-benefit (economic trade-off) analysis approach is often alluded to by scholars from different disciplines when talking about personal data sharing online. For example, Nissenbaum states, "*Individuals may choose to pay the "informational price" of disclosing personal data to social media and search engines, because the immediate benefits outweigh the potential harms, the social cost to the individual of opting-out is too high and the "notice and consent" model gives the individual, at the very least, a semblance of control".* [271] Such cost-benefit type analysis is often accompanied by warnings about the complexity of consumers' privacy decision-making, which fall short of assumptions of traditional economic theories of rational behaviour, particularly in the long-run. O'Hara/Shadbolt describe this as "*the classic type of privacy problem*", one where humans find it hard "*to balance the tangible benefits and the intangible costs*" when sharing personal data.[272] Economists such as Varian (1996)[273] articulate particular economic concerns around secondary usage of personal data whereby a consumer may have little knowledge of (or, indeed, control over) how an organisation will later use that data, even if the initial sharing of it was rational. An especially important risk factor in many data sharing scenarios is the risk of later re-identification from data that has been subject to anonymisation.

Such warnings have led to the formulation of more nuanced and granular views of the trade-offs associated with privacy protection and data sharing. For example, empirical studies of privacy valuations have been carried out, alongside studies into the degree of cultural basis in such privacy valuations. Such inquiries may help contribute to our understandings of how privacy as an abstract good (and personal information in its commodity form) is subjected to market forces by analysing the individual and social costs and benefits associated with its disclosure.

In brief, it is clear that the commercial interests that facilitate information flows are often underpinned by consumer-facing demand-led use. A shift in emphasis in the prevailing EU narrative to focus on effects-based approach (and mitigating harm) as consumer protection – and enforcement of consumer protection, with privacy/data protection branded as a consumer right, not just a human right - may be useful. At the same time, it is acknowledged that market forces (demand or supply driven) may offer incentives for compliance that are as sustainable as legal-led ones, if not more so.

[271] Nissenbaum (2011), p. 35.
[272] O'Hara & Shadbolt (2008), p.5
[273] Varian (2002).

Switching back to law, a critical analysis of the main legal scholars who are proponents of effects-based theories – or who propose theoretical variants to this approach – is set out in Section 4.3 below.

4.2 Legal interpretation and policy consistent with the Effects-based Approach

While there are no legislative provisions in the DPD (or DPA) providing explicit support for the effect-based approach, it is arguable that some provisions contain, implicitly, an effects-based emphasis. A general example in the DPD is reference to "*processing operations likely to present specific risks to the rights and freedoms of data subjects*".[274] Such references imply a focus of regulatory concern on potential negative impact associated with future processing operations and anticipative measures that can be taken in view of these. However, *"risks to the rights and freedoms of data subjects*" is not fully explained - risks of what exactly? This phrase merits further investigation.[275] Other legislative-provision examples with similar implications are found specifically in a personal data definitional context, as well as regarding other special categories of data, mentioned in the DPD/DPA and in the GDPR.

This section considers these examples and related policy guidance for the interpretation of the legislative definitions of personal data in the EU/UK. The following sub-sections discuss three inter-related theoretical aspects associated with determining when, and to what extent, data protection law applies in practice:

1) **Dissecting the concept of sensitive personal data from an effects-based perspective;**
2) **Considerations of processing-usage within the concept of personal data; and,**
3) **Other effects-based discourse used in interpreting the personal data concept and more generally relevant under data protection law.**

[274] Article 20 states: "*1. Member States shall determine the processing operations **likely to present specific risks to the rights and freedoms of data subjects and shall check that these processing operations are examined prior to the start thereof**. 2. Such prior checks shall be carried out by the supervisory authority following receipt of a notification from the controller or by the data protection official, who, in cases of doubt, must consult the supervisory authority. 3. Member States may also carry out such checks in the context of preparation either of a measure of the national parliament or of a measure based on such a legislative measure, which define the nature of the processing and lay down appropriate safeguards."* (emphasis added)

[275] See, e.g. Van Dijk et al (2016). This issue is revisited for further discussion later in this chapter and in Chapter 7 below.

4.2.1 Dissecting the concept of sensitive personal data from an effects-based perspective

The presence of the identificatory requirement, together with the other definitional requirements, might imply that all data satisfying such criteria are subject to the same degree of legal restriction. However, personal data processing is not the only determining factor regarding the extent to which data protection rules apply. Data protection law also embeds flexibility in respect of the type of rules that apply when a special category of sensitive personal data is processed.[276] Arguments are considered next for why the essence of the concept of sensitive personal data implies support for the Effects-based Approach, especially through its acknowledgement of (and value placed on) protecting against different degrees of harm associated with personal data processing in certain circumstances under data protection rules.

4.2.1.1 The DPD and its origins

Per Chapter 2, those who process sensitive personal data are subject to enhanced data protection obligations. While the DPD does not actually include the phrase 'sensitive personal data', Article 8 (headed "special categories of processing") includes an itemised list of data to which such special rules apply (s.1): information revelatory of racial or ethnic origin; political opinions; religious or philosophical beliefs; trade-union membership; and, data concerning a person's health or sex life. Article 8(2) DPD prohibits sensitive data processing unless specific-listed exceptions apply, e.g. the data subject gives their explicit consent to the processing activity.[277] The DPA adopts the same-style 'general prohibition followed by exceptions' legal approach to data deemed sensitive, with the special reasons that can justify sensitive data processing set out in its Schedule 3.[278]

To understand the logic behind the DPD's inclusion of the sensitive personal data category, pre-DPD differences in EU countries' approaches to certain types of data the processing of which were deemed particularly risky to data subjects are illuminating. As Wong explains, "*[c]ountries, like*

---

[276] As Solove/Schwartz point out (2013, #1, p.913): *"the concept of sensitive data shows how the European Union already supports different categories of data with different levels of protection. The Directive identifies a special category of data called "sensitive data" and provides it with stronger protections than other types of data"*). They go on, "*[t]hus, the EU approach already diverges from uniformity when different levels of protection will better protect individuals' right to privacy…"*

[277] Other exceptions that justify the processing of sensitive personal data include where the data subject has made the information public, or where there is a need to use such data in the establishment, exercise, or defence of legal claims.

[278] Compare, section 2, DPA: "*"sensitive personal data" means personal data consisting of information as to— (a)the racial or ethnic origin of the data subject, (b)his political opinions, (c)his religious beliefs or other beliefs of a similar nature, (d)whether he is a member of a trade union …(e) his physical or mental health or condition, (f)his sexual life, (g)the commission or alleged commission by him of any offence, or (h)any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings".* Of note, Article 8(5) DPA also makes special provision for data about criminal records: "*[p]rocessing of data relating to offences, criminal convictions or security measures may be carried out only under the control of official authority, or if suitable specific safeguards are provided under national law, subject to derogations which may be granted by the Member State under national provisions providing suitable safeguards".*

Austria and Germany…had consistently rejected all abstract categorisations of personal data and instead focussed on a context-orientated appreciation of the data…"[279] With the official introduction of the concept of a per se sensitive personal data classification in the Council of Europe's data protection Convention 108[280] that approach was swept-aside. According to Simitis, that decision "*sanctioned the quest for a particular regulatory regime of sensitive data, a position that since then has been over and over underscored by the Council's recommendations"*.[281]

Delving deeper into this legislative shift-change, Simitis queried in 1999 whether data sensitivity really was a valid classificatory criterion for determining processing requirements in the context of Convention 108.[282] He argued that that the list of sensitive personal data types enumerated therein should not be understood as definitive – i.e. a definitely closed list – but typifying only. For Simitis, sensitive personal data is merely an "*ephemeral indication*"[283] (a notional 'alarm bell') that "*signals*

---

[279] Wong (2007, p.13). See also Simitis (1999, p.1): "*[t]he early Norwegian attempts to elaborate methods permitting to distinguish personal data according to their sensitivity are as significant for the importance attached to a series of data whose processing was deemed to be particularly risky for the persons concerned as the clearly articulated demand of the French legislators to simply prohibit the use of such data… the initial discussions were … first and foremost debates on whether "sensitivity" really is a valid criterion for determining the conditions of the processing*". Indeed, at MS level, the original enumeration in the DPD was amended by various countries in adding to the list. See also Simitis (1999, p.3) discussing results of a survey about EU MS implementation of Convention 108: "*[t]hat the legislative intervention is all but a theoretical means to question and review the composition of the list is demonstrated by the answers to the questionnaire. Thus Finland amended the original enumeration in order to include trade union membership, an extension also envisaged by the Netherlands and Norway. Both countries intend besides to revise the actual enumeration, the first by striking out psychological data, the latter by redefining the reference to family affairs. Portugal, finally, renounced, as did Estonia, a special protection of data related to the property and to the financial situation of the data subjects, but included genetic data in its list".* See also p.7: *"[a]lready three of the laws that at least "in principle" favour uniform rules tolerate distinctions. The Austrian, French and Italian answers point to the high degree of sensibility of genetic data and the ensuing necessity to secure an equally higher protection. The Hungarian answer goes further and openly advocates a split. Especially the data concerning racial or ethnic origin, political opinions, party affiliations or religious beliefs are deemed to be more sensitive. Danish law intensifies also the protection of data related to "political matters", but only partially. As long as they have not been accessible to the general public their processing is prohibited".* See also Wong, ibid, p.10: *""[t]o give an example, the Council of Europe report (entitled Informational self-determination in the internet era) recommended that identification numbers that enable many databases or data to be connected together should be included within the definition of "sensitive data". This practice has become widespread in the public and private sector. The DPD however, leaves it to the discretion of the member state to determine the conditions under which a national identification number or any other identifier of general application may be processed (Art. 8(7) DPD), but does not specifically touch on the subject of identification numbers in the online environment or its use in databases".*

[280] Article 6 of Convention 108 provides that personal data revealing racial origin, political opinions or religious or other beliefs, as well as personal data concerning health or sexual life, may not be processed automatically unless domestic law provides appropriate safeguards. The same prohibition applies to personal data relating to criminal convictions. To note, however, the DPD differs from the Convention's approach in two main respects: the DPD includes the trade union membership as a specific category of sensitive data; and, whereas the DPD's list of sensitive personal data types is often considered exhaustive, the Convention's list is merely indicative. Notwithstanding, as discussed next, academics have queried whether sub-categories of sensitive personal data can in fact be implied anew under the DPA and the DPD. See, e.g. Wong (2007, p.9): *"[f]ollowing the Lindqvist case (C-101-01), it is questionable how the criterion applies in practice. More specifically, it can be contended that any images/photographs of the data subject uploaded on the internet falls within Art. 8 of the DPD because the image/picture reveals some of the characteristics that may be regarded as sensitive data".* Compare Elliot et al (2016, p.82): *"[w]ithin the DPA, the definition of 'sensitive' is based on a list of categories which are widely regarded to be an incomplete list of what might intuitively be covered (financial data for example is notably absent). So you may consider that there are other categories of data which are sensitive and identifying".*

[281] Simitis (1999, p.1).

[282] Ibid.

[283] Ibid, p.3 (in full): *"[t]he large majority of the actual laws may certainly suggest that the attribute "sensitive" is reserved to an exclusive class of data carefully selected by the legislators. None of these laws contents however itself*

that the rules normally applicable to the processing of personal data may not secure adequate protection" in particular situations.[284] For that reason, he argues that there can be no exhaustive list of sensitive data, nor can there be an unconditional prohibition on their processing-usage.[285] Instead, any personal data can be sensitive, given a particular processing-usage context adjudged as such, taking into account a range of context-specific factors, which assimilated "*allow both the range and the effects of the processing to be discerned and thus to determine its sensitivity*".[286] Therefore, he proposes that the legal test for determining whether sensitive personal data exists under the DPD should use a contextualised approach.[287]

4.2.1.1.1 Regulatory guidance

The WP

In its Statement on the role of a risk-based approach in data protection legal frameworks (WP218)[288], the WP acknowledges that data's sensitivity can provide different threat degrees to a subject's rights/freedoms, including regarding privacy risks that might flow from its processing:

---

with the statement that its list is exhaustive. On the contrary, they all provide ways and means to reopen the apparently definitely closed list. *The proviso of a legislative intervention [e.g. Estonia] … demonstrates that there is no definitive list of sensitive data…. Each of these cases demonstrates that the enumeration of sensitive data is throughout understood as an ephemeral indication".*

[284] Ibid, p.8. The quote goes on: *"[b]oth the starting point and the range of all considerations are determined by the potential contexts of the processing. They permit the specific risks to be discerned and the antidotes to be designed. Prohibition is hence a possible but by no means a compelling consequence. And even where it appears justified to forbid the use of certain data, the prohibition remains a reaction confined to the context that legitimates and at the same time limits the exclusion of the processing".*

[285] Ibid. In other words, for Simitis (p.8), this would explain why there developed a "special regime" for all data deemed sensitive, but it is also a regime under which processing-oriented considerations prevail as "***their use*** *may be generally regarded as a possible source of particular risks for the data subjects. However, whether and to what extent these risks justify an exclusion of their processing, is a question that can only be answered separately for each of these data and in consideration of the circumstances characteristic of the specific use"* (emphasis added).

[286] Ibid, p.5. Simitis reasoned that it is vital to consider contextual factors when determining the sensitivity of data, including the interests of the data controller, as well as the potential recipients of the data, the aims for which the data are collected, the conditions of the processing, as well as its possible consequences for the individual and others. Also see, from p.5: *"[s]ensitivity is no more perceived as an a priori given attribute. On the contrary, any personal datum can, depending on the purpose or the circumstances of the processing be sensitive. All data must consequently be assessed against the background of the context that determines their use. The specific interests of the controller as well as of the potential recipients of the data, the aims for which the data are collected, the conditions of the processing and its possible consequences for the persons concerned are factors that, put together, allow both the range and the effects of the processing to be discerned and thus to determine its degree of sensitivity. An evaluation of the sensitivity requires hence more than a mere look at the data. It may very well be that, for instance in the case of genetic data or of data concerning criminal convictions, the risks for the data subjects are more or less obvious. However, the sensitivity can in the end only be affirmed if all the elements typical of the particular processing operation are taken into account. …A clearly limited use for purposes exhaustively defined by law and a deliberate exclusion of numbers providing information on their own on the data subjects are…amongst the measures that allow, but at the same time reduce, the use of personal identificators to a few precisely indicated processing operations under clearly prescribed conditions. In sum, absolute classifications of personal data are here as elsewhere supplanted by a distinctly situational assessment. Sensitivity is no longer an attribute granted once and for all but a characteristic determined by the context of the intended use that therefore has in principle to be constantly reappraised".*

[287] Ibid, p.9.

[288] Article 29 Working Party (2014, WP218).

> The legal regime applicable to the processing of special categories of data … can also be considered as the application of a risk-based approach: strengthened obligations result from processing which is considered risky for the persons concerned.[289]

A bit later, the WP includes a similar statement:

> Risks, which are **related to potential negative impact on the data subject's rights, freedoms and interests**, should be determined taking into consideration specific objective criteria such as the nature of personal data (**e.g. sensitive or not**), the category of data subject (e.g. minor or not), the number of data subjects affected, and the purpose of the processing. The severity and the likelihood of the impacts on rights and freedoms of the data subject constitute elements to take into consideration to evaluate the risks for individual's privacy...[290] (emphasis added)

This quote hints at possible support for an effects-based perspective in respect of the concept of sensitive personal data: the latter is assumed to indicate a factor denoting the possibility of "*negative impact on the data subject's rights, freedoms, and interests*". Notwithstanding, the WP does not specify here what that phrase means exactly; does it require a likely (concretised) effect?[291] (For more consideration of this phrase, see Section 4.4 below).

The ICO

Support for an effects-centric perspective on sensitive personal data is also visible in some ICO statements. In its European Commission consultation-response on the legal framework for the fundamental right to protection of personal data, for example, it promotes an effects-based approach of sorts over a rigid categorisation of sensitive personal data categories:

> The Information Commissioner suggests a definition stating that information is sensitive if its processing would have an especially adverse or discriminatory effect on particular individuals, groups of individuals or on society more widely. This approach allows for flexibility in different contexts so that real protection is given where it matters most.[292]

---

[289] Ibid, p.2.
[290] Ibid, p.4.
[291] Compare, ibid, p.4: *"[i]n the context referred to above, the scope of "the rights and freedoms" of the data subjects primarily concerns the right to privacy but may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion".*
[292] IC (2009), p.3 (in full): *"the rigid categorisation of special categories of data is not an effective way to allow acceptable processing but prohibit the unacceptable. A more flexible and contextual approach is needed.* ***The Information Commissioner suggests a definition stating that information is sensitive if its processing would have an especially adverse or discriminatory effect on particular individuals, groups of individuals or on society more widely. This approach allows for flexibility in different contexts so that real protection is given where it matters most.*** *In practice, it could mean that the current list of special data categories remains largely valid, but it would allow for*

4.2.1.1.2 The GDPR

Like the DPD, the GDPR does not refer to sensitive personal data explicitly in its articles, although it does refer to special categories of personal data at Recital 53 that "*merit higher protection*".[293] Moreover, Article 9 entitled "*processing of special categories of personal data*" adds new legal bases (processing grounds) for processing sensitive personal data.[294]

The GDPR introduces three new types of sensitive personal data in line with 21st century concerns: "*genetic data*", "*biometric data for the purpose of uniquely identifying a natural person*", and data about sexual orientation.[295] These additions illustrate a development in the debate over the relativity of the sensitive data list.[296] Specifically, it queries whether the DPD's intention was for Article 8(1) to be exhaustive or non-exhaustive. Arguably, the GDPR additions demonstrate the

---

personal data not currently in the list to be better protected, for example financial data or location data. Or, more radically, the distinctions between special categories and ordinary data could be removed from the new framework, with emphasis instead on the risk that particular processing poses in particular circumstances" (emphasis added)**.** Later on in the same guidance, (p.12): *"[i]t is important to give a message to data controllers that a simply binary (special categories – the rest) approach is not good enough,* ***and they must consider the context in which they hold information and the risk this poses to individuals.*** *It would be helpful if national data protection authorities or EU-level bodies, such as the Article 29 WP, could produce guidance with examples that could help organisations to assess genuine sensitivity in various contexts"* (emphasis added). Indeed, this quote hints at possible direct support for the Effects-based Approach as it concerns the concept of personal data, not just sensitive personal data, more generally in law. See also IC, (2010, p.10). In that respect, compare also the assertion by Hon et al (2011, at p.227) suggesting that the ICO may be seen as showing support for a similar risk-based approach to personal data generally, and not just to sensitive personal data referring to this quote: *"[t]he ICO has pointed out that a fixed list of categories can be problematic: sensitivity can be subjective/cultural, a set list does not take account sufficiently of the context and may even exclude data which individuals consider sensitive, and non-EU jurisdictions may have different lists, which could cause difficulties for multinationals".*

[293] Recital 10 GDPR also explicitly mentions the term 'sensitive data' with reference to such special categories of personal data: "*[t]his Regulation also provides a margin of manoeuvre for Member States to specify its rules, including for the processing of special categories of personal data ('sensitive data'). To that extent, this Regulation does not exclude Member State law that sets out the circumstances for specific processing situations, including determining more precisely the conditions under which the processing of personal data is lawful*".

[294] Article 9(2)) GDPR refers to the processing of such categories being deemed necessary solely for the following, additional (alternate) reasons: for substantial public interest (g); for individual health purposes (h); for public health reasons (i); or, for archiving, scientific, historical, or statistical purposes linked to the public interest (j).The GDPR provides additional safeguards in connection with the processing of data relating to criminal convictions and offences, as well as processing for historical, statistical, and scientific research purposes. MSs are also free to adopt further safeguards for the processing of genetic, biometric, and health data.

[295] Article 9(1) GDPR: *"[p]rocessing of special categories of personal data 1.Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of* ***genetic data, biometric data for the purpose of uniquely identifying a natural person****, data concerning health or data concerning a natural person's sex life* ***or sexual orientation*** *shall be prohibited"* (emphasis added). See also Article 9(2-4) GDPR. To note, Article 4(15) GDPR defines 'data concerning health' anew and broadly as, "*personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status".* However, the definition of 'biometric data' is narrowly drawn in yet more ambiguous identification terms in Article 4(14) as meaning, "*personal data* ***resulting from specific technical processing*** *relating to the physical, physiological or behavioural characteristics of a natural person,* ***which allow or confirm the unique identification*** *of that natural person, such as facial images or dactyloscopic data*" (emphasis added). Compare, in this respect, the definition of 'genetic data' in Article 4(13) as meaning, *"personal data relating to the inherited or acquired genetic characteristics of a natural person* ***which give unique information about*** *the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question*" (emphasis added).

[296] As referred to above, indeed, at MS level, the original enumeration in the DPD was amended by various countries in adding to the list.

impossibility of a once-and-for-all-time definitive enumeration of members of such lists, including data types the processing of which are done in order to uniquely identify natural persons.[297] Furthermore, such lists require reviewing periodically.[298]

What about effects? Recital 51 GDPR explains the rationale for stricter processing rules for sensitive data:

> Personal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms merit specific protection **as the context of their processing could create significant risks to the fundamental rights and freedoms**. (emphasis added)

Said otherwise, the GDPR suggests that the processing of sensitive personal data can be a possible source of significant risks to data subjects. Does this statement support the use of an effects-based approach perspective in categorising sensitive personal data? Problematically, the GDPR does not specify "*risks to fundamental rights and freedoms*" in what ways – i.e. it remains unclear to what extent this phrase refers to an effects-based interpretation (requiring negative impact potentially suffered in concreto). However, in minimum, by including the word "significant" this quote introduces the notion of risk quantifiability (the risk of something bad occurring – whatever that may be – as capable of estimation in terms of degree of likelihood and/or magnitude of impact) as a relevant legal consideration. Similarly, assessing risk likelihood and effect magnitude is also a feature of the effect-based approach (see Section 4.4).

4.2.1.1.3 Summary

As highlighted in the UK Anonymisation Decision-making Framework (discussed further in Chapter 5), the data sensitivity concept leads us into "*thinking about the impact side of risk*"[299] and its connection with potential harm ("*underlying the notion of sensitivity is one of potential harm*").[300]

---

[297] McCullagh (2007, p.193): "*[i]t is important to review the continuing relevance of existing categories of sensitive data in the Directive in the light of changes in societal structures and advances in technology. In the pre-computer era, data processing was not automatic and large-scale, uncontrolled surveillance was costly, thus providing natural barriers for privacy protection".* See also p.200: *"[i]n the 21st century, new concerns have risen; for example, developments in the fields IT and biometrics are raising new potential categories of sensitive data. Indeed, findings from interviews and the survey indicate that whilst not all of the legally recognised categories of data continue to be perceived as sensitive, some which are not legally recognised categories of data are emerging which are considered extremely sensitive."*

[298] Ibid, p.190: "*[t]he concept of 'sensitive' data was first considered for introduction into international law by the expert group drafting the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980). Sweden and the German state of Hesse had already incorporated the concept into national and state law.* ***Ultimately the drafters of the Guidelines decided not to include extra safeguards for designated categories of sensitive data. The absence of safeguards seems to be partly due to a failure to achieve consensus on which categories of data deserve special protection, as the guidelines state: "...it is probably not possible to define a set of data which are universally regarded as being sensitive. (para 19 (a))""*** (emphasis added).

[299] Elliot et al (2016, p.82).

[300] Ibid, p.65. The document (p.81) goes on to connect the notion of increased re-identification risk and the likelihood of negative effects occurring following the disclosure of anonymised data (discussed further in Chapter 5): "*[s]ensitive*

However, references to sensitivity here tie – not just to consideration of the type of data present - but also the proposed processing-usage context. So, while "*one's address is for most people a fairly innocuous piece of information but for somebody on a witness protection scheme it becomes highly sensitive*".[301] Said otherwise, connection is made between data sensitivity and the potential for a piece of information to cause harm to the individual to whom it relates when processed in particular circumstances. This argument stands up in practice. While the sensitivity of some types of data, such as 'genetic fingerprints', are indisputable, in reality, examples of data not typically considered sensitive can illustrate how context-oriented, but also use-dependent, considerations affect perceptions of data sensitivity. McCullagh similarly argues for scoping an approach to the legal concept of sensitive personal data considering also the data holder's intentions behind its processing (see the next sub-section (4.2.2)).

For thesis purposes, proffering a definitive opinion on whether to do away with the category of sensitive personal data altogether - in favour of a contextualised use-based model to data sensitivity - is obiter to answering the research questions. Regardless, under existing data protection law, the rules on sensitive personal data do not take account of data sensitivity in a usage context; what matters is whether the information concerned falls within the enumerated data category. Thus, data types falling within this delineated category are automatically subject to stricter rules when processed whatever the processing-usage context.[302] For example, processing information about an individual's health (e.g., whether they have a cold, or are HIV-positive) requires similar treatment, including justification under a legal basis (satisfying Article 8(2) DPD, or Schedule 3 DPA), irrespective of processing context.

Notwithstanding, and by analogy to Simitis' arguments, the identificatory requirement may also be seen as merely an 'alarm device'. Its role is to incite ex-ante reflection upon the potential negative effects of a processing activity involving data relating to a particular person upon that individual. While it signals that the application of data protection rules to the processing of data relating to specific individuals is justified, arguably, its ultimate goal became clouded and its value diminished

---

data is thought to increase re-identification risk because (i) it is more likely to be targeted because it is interesting, ***and (ii) the impact (and potential harm) of a disclosure may be greater***" (emphasis added).

[301] Ibid, p.82.

[302] Whereas Wong (2007, p.14) refers to the argument that "*sensitivity is determined according to its context and not simply based on an enumerative list under Art. 8(1)".* Wong then goes on to argue (p.12) that the "*current categorisation (under Art. 8 DPD) based on the definition of the actual nature of data appears impractical and arguably antiquated with the general processing of personal data on the internet".* See also Hon et al (2011, p.227): "*the distinction between personal data and sensitive data is, arguably, also no longer tenable. The European Commission is to consider whether other categories of data should be considered as 'sensitive data', for example, genetic data; and will be further 'clarifying and harmonising' conditions under which sensitive data may be processed. The UK ICO has made the point that a fixed list of categories can be problematic: sensitivity can be subjective/cultural, a set list does not take account sufficiently of the context and may even exclude data which individuals consider sensitive, and non-EU jurisdictions may have different lists, which could cause difficulties for multinationals".*

as an instrumental concept. However, the underlying concern signalled by this alarm remains, which regards the harmful effects that might otherwise result from specific processing activities in context. The time has come to make this more explicit.

4.2.2 Considerations of processing-usage intrinsic to the concept of personal data

This sub-section provides support for the existence of a usage-oriented viewpoint upon the legal concept of personal data under existing law, by analogy giving support to the Effects-based Approach. Such viewpoint involves consideration of the processing circumstances planned in respect of data relating to a particular individual in defining whether it is personal or not.[303] Variants upon this perspective include the (so-called) 'purpose-based approach' to personal data.[304]

Under existing interpretations of the identificatory requirement, per Chapter 3, there is already support for purpose-based considerations in defining personal data. Where the intention behind data processing is to reveal the identity of the individual to whom the data relates, the WP[305] and ICO[306] have held this to be relevant for determining whether the identificatory requirement is met

---

[303] In other words, it is a dynamic model that transcends static considerations around the nature of data alone.

[304] As the name suggests, it focuses upon the purpose of the data holder in processing information relating to a particular person in a particular context as determinative of whether that information is personal data in law. (To note, this is a slightly different from the contextualised approach discussed in the last sub-section, for example, the approach argued by Simitis (1999) that personal data becomes sensitive according to its context, albeit related). By way of comparison, McCullagh (2009, at p.199) gives an example of what a purpose-based approach means in relation to the sub-concept of sensitive personal data: "*[t]he Council of Europe (2005) proposed a purpose-based approach which would consider the purpose underlying the processing of personal data, that is, whether the processing is intended to reveal sensitive data. "This approach would make it possible to consider the actual processing of data as sensitive rather that the data itself, even if no sensitive data were involved. For example, a search of trips to Rome conducted by a web surfer using Google or his or her purchases of religious books, reading of a papal encyclical, etc, may be treated as revealing a religious opinion". (Poullet et al 2004) Searching for information on a trip to Rome would not in itself constitute processing of sensitive religious information, but when it is combined with searches for Vatican city visiting hours the purpose of the information processing may change. Of course, searches for such information may be purely coincidental, for instance if a person has heard that a restaurant within the Vatican grounds is worth a visit and checks the opening times etc"".* Compare, Wong (2007, p.10) - referring to the Council of Europe-commissioned report by Dinant & Poullet (2004) - "*[i]t is interesting to note from the report that the current definition of sensitive data is too wide and we should abandon the approach based on the definition of the actual nature of data in favour of a purpose-based approach. To put it another way, what is the purpose of such processing? Is the processing intended to reveal sensitive data such as political opinions? This alternative would not only be pragmatic, but also resolve the difficulties highlighted in Lindqvist, where any personal information published on the web could theoretically fall within Art. 8(1) involving the processing of sensitive data".*

[305] See WP 136, p.15, discussing Recital 26 of the DPD and means to identification: "*[t]he cost of conducting identification is one factor, but not the only one.* ***The intended purpose****, the way the processing is structured, the advantage expected by the controller, the interests at stake for the individuals, as well as the risk of organisational dysfunctions (e.g. breaches of confidentiality duties) and technical failures* ***should all be taken into account*****"** (emphasis added). See also p.19: **"***the identification of individuals (to apply the appropriate treatment in case of need) is one of the purposes of the processing of the key-coded data… The identification of patients is thus embedded in the purposes and the means of the processing";* and p.16 (which gives examples of the importance of data controller purpose-intent to determining personal data in relation to video-surveillance, dynamic IP addresses, and damage caused by graffiti).

[306] For example, for further discussion in Chapter 5, see its Anonymisation Code of Practice (2012, p.23): *"[c]learly, some sorts of data will be more attractive to a 'motivated intruder' than others. Obvious sources of attraction to an intruder might include: for nefarious personal reasons or financial gain; the possibility of causing mischief by embarrassing others; revealing newsworthy information about public figures; political or activistic purposes, eg as part of a campaign*

(under the Means Test). Moreover, in the AG Opinion in the *Nowak* case, AG Kokott placed emphasis on a purposive approach in a more fundamental sense (relative to the overall purpose of the DPD in protecting an individual's private life) in defining personal data:

> [A]n examination script incorporates information about the examination candidate and is in that sense a collection of personal data. That this is the correct conclusion is also shown, moreover, in the fact that an examination candidate has a legitimate interest, based on the protection of his private life, in being able to object to the processing outside the examination procedure of the examination script ascribed to him. An examination candidate does not have to accept that his script can be disclosed to third parties or published without his permission.[307]

Analogously, under the Effects-based Approach, considering the purpose of a particular processing activity with respect to a particular piece of information might be deemed a relevant factor in determining whether it is likely to have a negative effect upon an individual to whom it relates. For example, is the intention behind the activity to harm the individual to whom the data relates? If so, this might suggest that harm would indeed be likely to flow from the activity once carried out.

Upon reflection, however, it is clear that a usage-oriented perspective to the legal concept of personal data can embrace more than just purpose (data holder intention) considerations. For example, consider an ex-ante assessment of the question, 'is the processing of this particular bit of information in a particular context likely to be used to harm the individual to whom the information relates?'[308] To assess this, it seems relevant to take also into account other case-relevant circumstances, including likely result considerations associated with data processing-usage beyond the intentions of the one doing the data processing.

4.2.2.1 Regulatory guidance

4.2.2.1.1 The WP

There is already a way in which the WP adopts a usage-oriented perspective in providing guidance on how to assess whether data are personal data in law.

---

against a particular organisation or person; or curiosity, eg a local person's desire to find out who has been involved in an incident shown on a crime map".

[307] C-434/16, Nowak v. Data Protection Commissioner, Opinion of Advocate General Kokott delivered on 20 July 2017, paras 25-26.

[308] Compare the ICO's Anonmyisation Code of Practice (2012, p.23): "*[o]ne example might be health data, where, although there may be no obvious motivation for trying to identify the individual that a particular patient episode relates to, the degree of embarrassment or anxiety that re-identification could cause could be very high*".

First, and separate from its discussion about the identificatory requirement, the WP highlights the importance of considering intention behind data usage in relation to a specific processing activity as a relevant factor in assessing whether data *relates to* a particular individual. In WP136, per Chapter 2, the WP analyses the criteria that compose the concept of personal data and adopts a wide interpretation, particularly on the question of when data may be deemed 'relating to' an individual. The opinion provides three *alternative* elements – i.e. 'content', 'purpose', or 'result' – to determine whether information relates to an individual (i.e. as part of its definitional ambit). With respect to the 'purpose' element', it then describes this as follows:

> That "purpose" element can be considered to exist when the 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.[309]

The WP has also made a connection between the purpose behind data usage and the definitional nature of data in another legal sense. In 2015, within an annex to a letter addressed to the Director of Sustainable and Secure Society at the European Commission,[310] the WP identifies relevant criteria for determining when data processed by lifestyle and wellbeing apps and devices should be considered 'health data' (legally defined).[311] Highlighted is the criterion of data usage, specifically processing purpose. For example, the WP opines that even if data under consideration does not seem to be health data on the face of it, it can still fall within the category of health data if it is collected in order to determine the health status of data subjects. Moreover, the WP states that health data:

> [M]ay also include cases where a controller uses any personal data (health data or not) with the purpose of identifying disease risk (such as, for example, investigating exercise habits or diet with the view of testing new, previously unknown or unproven correlations between certain lifestyle factors and diseases).[312]

---

[309] Ibid, p.10.

[310] Article 29 Working Party. Letter to Mr Paul TIMMERS Director of Sustainable and Secure Society Directorate DG Connect, Brussels, 25 February 2015.

[311] Ibid, Appendix. For background, the WP acknowledged that the DPD does not clearly define what falls within the category of health data within the meaning of its Article 8 (regarding sensitive personal data). In effect, ***"defining the category of health data is important to determine in what circumstances the data processed by lifestyle and wellbeing apps and devices are to be considered data about health"***.

[312] Ibid, p.3. For example, the WP refers to "sad" messages sent by users, opining that when examined "*for the purpose of diagnosis/health risk prevention or medical research*" such messages constitute health data. This has led Stalla-Bourdillon (2015, The Article 29 Working Party on the concept of health data: could it mean that we need to adapt the definition of health data as well as that of personal data?, [online]) to conclude as follows: *"when raw data are collected for the purpose of describing or expressing the health status of a data subject, the whole dataset become health data…In consequence, what seems to be the crucial criterion is* ***the usage of data*** *and not so much its very nature,* ***which could mean that in the end the criterion is more subjective than objective: what matters would be the intention of the***

By analogy, these quotes provide support for applying the same kind of reasoning to determine the remit of the category of personal data itself – in lieu of the identificatory requirement - under EU data protection rules via the introduction of the Effects-based Approach.[313]

A clue to what factors beyond purpose/intention may be considered relevant to such an assessment can also be seen in WP136 when discussing the aforementioned 'result' element (of the 'relating to' criterion). That discussion shows clearly, once again, that the criterion of the usage of the data is considered already an important aspect of defining personal data legally:

> A third kind of 'relating' to specific persons arises when a **"result"** element is present. Despite the absence of a "content" or "purpose" element, data can be considered to "relate" to an individual because **their use 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 be a major impact. It is sufficient if the individual may be treated differently from other persons as a result of the processing of such data.[314] (emphasis added)

Said otherwise, the WP also interprets the 'result' element in a context-specific way.[315] It gives a scenario example where the 'result' element may be deemed present because the use of data is likely to have an impact on a person's rights/interests in a particular processing context.[316]

---

data controller rather the characteristics of the data themselves [unless an objective or eventually mixed standard is used to determine the intention of the data controller, e.g. that of a vigilant data controller; it could be argued in such a case that if a rich dataset is in fact collected, the intention of the data controller must have been that of characterising the health status of data subjects]" emphasis added.

[313] Bearing in mind, as mentioned, that the purpose behind a data processing activity is just one factor that may be relevant to assessing whether there might be a Relevant Effect likely occurring in respect of a person to whom (it is already established that) the data relates.

[314] WP 136, p.11. Compare also, p.25, which states that the 'relating to' element "*covers information that may have a clear impact on the way in which an individual is treated or evaluated*"**.** In that context, Zwenne (2013, p.6) argues that discussion of identifiability by the WP appears passed over: ***"****[t]he Working Party determined that there is already identifiability when 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". And further:* "[*i]t is sufficient if the individual may be treated differently from other persons as a result of the processing of such data". …In this interpretation of the concept of personal data, the Working Party ignores what I call the aspect of identity…contrary to what has sometimes been suggested, radical new interpretation of the concept. All of a sudden, for the Working Party it was no longer a matter ascertaining the identity of the individual involved"*. This perception by Zwenne may be linked to the fact that the legal significance of the potential for a certain individual to be treated differently because of the processing of data is interwoven by the WP in WP136 with reference to the capacity for "web traffic surveillance tools" – i.e. device "identifiers" - to be used on the Web to single someone out. See, WP136, p.14: *"the individual's personality is pieced together* ***in order to attribute certain decisions to him or her****. Without even enquiring about the name and address of the individual it is possible to categorise this person on the basis of socio-economic, psychological, philosophical or other criteria* ***and attribute certain decisions to him or her*** *since the individual's contact point (a computer) no longer necessarily requires the disclosure of his or her identity in the narrow sense"* (emphasis added). Notwithstanding, this thesis argues that the 'relating to' element and the identificatory element may be distinguished in discussions of their application to a particular set of facts.

[315] Compare also WP136, p. 12: *"[r]esulting from the previous analysis, the question of whether data relate to a certain person is something that has to be answered for each specific data item on its own merits".*

[316] Ibid, p.11 (see also fn.83 above): "*[e]xample No. 8: monitoring of taxis' position to optimize service having an impact on drivers. A system of satellite location is set up by a taxi company which makes it possible to determine the position of*

Similarities can be drawn here between the 'result' element and the Effects-based Approach, albeit that the latter requires a 'Relevant Effect' (see section 4.4 below) to be likely to occur following the processing of data (that would have already been deemed by the data holder as) relating to a particular individual. Nonetheless, such similarities illustrate a conceptual link by the WP between assessing data usage in particular circumstances and the resulting likely impact on particular individuals, on the one hand, and the transformative process by which data evolves into personal data under EU data protection law, on the other. Said otherwise, the WP appears to embrace a dynamic concept of personal data, with the corollary implication that the same piece of information may retain different characteristics legally at the same time, depending on the circumstances present in each circumstance.[317]

4.2.2.1.2 The ICO

The ICO has also highlighted the importance of assessing the purpose behind data processing, or likely effect following data processing, to determining whether data are personal. For example, in the TGN, in providing guidance on its interpretation of the 'relating to' criterion, the ICO states that answering positively to the following question means that the data is personal under the DPA - "*Is the data used, or is it to be used, to inform or influence actions or decisions affecting an identifiable individual?*"[318] On the same page, the ICO notes, "*context is important here*".[319]

Admittedly, focusing on the 'relating to' requirement,[320] as the ICO is doing here, is separate from focusing on the identificatory requirement also discussed in the TGN (or, indeed, the Effects-based Approach proposed to replace the latter requirement). Nevertheless, notably, the ICO appears to confuse its analysis of the two requirements in the TGN. For example, the ICO discussed how the same data may be personal data for one party but not for another party in the context of "*different*

---

available taxis in real time. The purpose of the processing is to provide better service and save fuel, by assigning to each client ordering a cab the car that is closest to the client's address. Strictly speaking the data needed for that system is data relating to cars, not about the drivers. ***The purpose of the processing is not to evaluate the performance of taxi drivers, for instance through the optimization of their itineraries. Yet, the system does allow monitoring the performance of taxi drivers and checking whether they respect speed limits, seek appropriate itineraries, are at the steering wheel or are resting outside, etc. It can therefore have a considerable impact on these individuals, and as such the data may be considered to also relate to natural persons. The processing should be subject to data protection rules****"* (emphasis added).

[317] Compare, WP136, p.11: "*[a] corollary of this is that the same piece of information may relate to different individuals at the same time, depending on what element is present with regard to each one*".

[318] TGN, p.12. It later uses similar – although not identical - language on this topic (pp.16-17): *"data may be personal data because it is clearly 'linked to' an individual because it is about his activities and* ***is processed with the purpose of determining*** *or influencing* ***the way in which that person is treated"****.* See also p.13: *"[a]s soon as data about a house is either: - linked to a particular individual, for example, to provide particular information about that individual (for example, his address)…; or - used in deliberations and decisions concerning an individual* ***(even without a link to the individual's name****, for example, the amount of electricity used at the house is used to determine the bill the individual householder is required to pay); then that data will be personal data"* (emphasis added).

[319] Ibid, p.12.

[320] Which, per Chapter 2, is arguably just about forming some nexus between data and a person but at a very low standard that could be accomplished in many different ways.

organisations processing the same data for different purposes".[321] It then goes on to illustrate this point using examples. The ICO describes two "*almost identical photographs*" taken of New Year celebration revellers in Trafalgar Square taken by two separate photographers and stored electronically on their computers (per fn.151 above):

> The first photographer, a photo journalist, takes a picture of the crowd scene to add to his photo library. The second photographer is a police officer taking photos of the crowd scene to identify potential troublemakers. The data in the electronic image taken by the journalist is unlikely to contain personal data about individuals in the crowd as it is not being processed to learn anything about an identifiable individual. However, the photo taken by the police officer may well contain personal data about individuals **as the photo is taken for the purpose of recording the actions of individuals who the police would seek to identify, if there is any trouble, so they can take action against them…**[322] (emphasis added)

By comparison, in earlier guidance the ICO seems to equate an intention to do certain things through processing web data (e.g. to use such data to make deliberations and decisions concerning nameless individuals) with the concept of an intention to single them out (and, therefore, per Chapter 3, an intention to identify them):

> If the information about a particular web user is built up over a period of time, perhaps through the use of tracking technology, **with the intention that it may later be linked to a name and address, that information is personal data**. Information may be compiled about a particular web user, but there might not be any **intention** of linking it to a name and address or e-mail address. There might merely be an **intention** t**o target that**

---

[321] Ibid, p.14:*"**[d]ifferent organisations processing the same data for different purposes** - It is important to remember that the same piece of data may be personal data in one party's hands while it may not be personal data in another party's hands… A single piece of data, which is not personal data for one data controller may become personal data when it is passed to another data controller".* At page 16, the ICO concludes, "*data may not be personal data in the hands of one data controller (for example, the estate agent) but the same data may be personal data in the hands of another data controller (for example, the police) depending on the purpose of the processing and the potential impact of the processing on individuals*".

[322] Ibid, p.15. A second example (as also alluded to in fn.151 above) is given by the ICO of an estate agent taking a photo of a high street shop to the market the property (for holding in digital form), which captures images of pedestrians walking past the shop when the photo was taken (ibid, pp.15-16): "*[t]he estate agent is not processing the shop data to learn anything about any of the pedestrians whose images were captured by chance on the photo, nor is it likely that the estate agent would ever process the photo for that purpose. **The estate agent is unlikely to possess the appropriate software to digitally enhance the photo to identify individuals**. Therefore, in the hands of the estate agent, the photo does not contain personal data about the pedestrians as it is not processed to learn something about those individuals and nor is it likely to be processed by the estate agent for this purpose….The estate agent might supply the police with a copy of the photo in response to the appeal. **The police would then process the digital photo, not to learn anything about the shop but, using photo enhancing technologies, to attempt to identify potential witnesses or suspects**. The photo would then be being processed to learn something about the individual pedestrians and, in the hands of the police, may be personal data about such individuals*" (emphasis added).

> **particular user with advertising, or to offer discounts** when they re-visit a particular web site, on the basis of the profile built up, without any ability to locate that user in the physical world. **The Commissioner takes the view that such information is, nevertheless, personal data. In the context of the on-line world the information that identifies an individual is that which uniquely locates him in that world, by distinguishing him from others.** [323] (emphasis added)

It is not clear whether the ICO is also referring to the 'relating to' criterion and its application in this quote.

Thus, there appears to be some confusion for the ICO (as well as the WP) about why finding out the purpose behind data processing is important in determining whether data are personal or not. Notwithstanding, as with the WP, the above quotes illustrate that the ICO also sees the personal data concept as dynamic involving consideration of critical factors detached from the data's nature alone, rather associated with the processing-usage context. This provides additional support for the doctrinal 'pedigree' of the Effects-based Approach as a proposed replacement for the identificatory requirement, which too relies on a dynamic concept of personal data liable to flux. It also requires consideration of the processing-usage context as part of reaching a conclusion about the likely effects of carrying out a particular processing activity upon data (that would have already been established as) 'relating to' a particular person.

4.2.2.2 The GDPR

Recital 76 GDPR states broadly, "*[t]he likelihood and severity of the risk to the rights and freedoms of the data subject should be determined by reference to the nature, scope, context* ***and purposes of the processing****"* (emphasis added). In trying to understand how this theme of risk-assessment related considerations of processing-usage fits into the GDPR's discussion of the personal data concept, Recital 30 GDPR is useful:

> Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, **may be used to create profiles of the natural persons** and identify them. (emphasis added)

[323] ICO (2001, p.12). Notably, there is not mention here of the Means Test by the ICO.

Thus, the GDPR suggest that those who carry out data processing activities to profile[324] individuals may find themselves subject to data protection rules *precisely because* such activities are likely to lead to the identificatory requirement being met as a matter of course (albeit there is no mention of Recital 26 GDPR here). Again, we see an emphasis upon processing-usage as a factor for consideration in deciding whether data are personal.

Recital 24 GDPR also associates the activity of profiling with intentions by profilers to take decisions about relevant individuals; 'natural persons' are "*tracked on the internet including potential subsequent use of personal data processing techniques…particularly in order to take decisions concerning him or her for analysing or predicting her or his personal preferences, behaviours and attitudes*". This suggests that profiling activity concerns - while associated with satisfaction of the identificatory requirement- actually belie concerns about the possibility of processing activities resulting in negative effects befalling singled-out relevant individual. In this way, it implies support for the Effects-based Approach in addressing such concerns head-on.

Article 22 and mirroring Recital 71 GDPR also draws comparisons.[325] They allude to serious data protection concerns arising potentially in an automated individual decision-making and profiling context where the activity "*produces legal effects concerning him or her or similarly significantly affects him or her*". Reference to "significantly" also hints at the importance of the magnitude of impact (that is likely) to flow from processing-usage as suggestive of whether the data being processed are deemed sufficiently worthy of being labelled personal data. It provides a clue to the delineation of 'Relevant Effect' under the Effects-based Approach for exploration. Otherwise, as Booth et al highlight: "*an inclusion within "personal data" of all data as potentially affecting an individual…could include almost any data" and "could lead to gross over-expansion of the legal framework*".[326]

---

[324]Article 4(4), GDPR: *"'profiling' means 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, in particular to analyse or predict aspects concerning that natural person's*** *performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements"* (emphasis added).

[325] Reflecting the content of Article 22(1), Recital 71 GDPR describes profiling as potentially illustrative of automated profiling based solely on which a "*data subject should have the right not to be subject to a decision…evaluating personal aspects relating to him or her…which produce legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention"*. Notwithstanding, this proviso is subject to the following caveat (reflected from Article 22(2)-(4)): "*decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent. In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision".*

[326] Booth et al (2004, p.18): "*[i]t means that information, ostensibly not about individuals, would come under full remit of data protection law based on a possibility of it being linked to individuals at some point in time*".

4.2.2.3 Summary

A usage-oriented perspective on the concept of personal data is compatible with a contextual perspective upon the same. Both appear endorsed in guidance and legislation to some extent, along with consideration of how processing activities carried out in particular circumstances may have possible consequences for the individual to whom that data relates.

As mentioned, the Effects-based Approach marries all three perspectives as, under it, the question needs addressing whether negative effects upon an individual are likely to arise from a particular processing activity. This, in turn, requires consideration of related factors of relevance, such as the purposes behind the processing.

4.2.3 Other effects-based analysis used in interpreting the personal data concept and more generally relevant under data protection law

Implicitly or explicitly, a good deal of EU data protection discourse already concerns the concept of risk. Per Chapter 3, the identification risk-based model has significant support from policymakers and scholars.[327] However, there is also implicit recognition that EU data protection law endorses the carrying out of effects-centric risk analyses in ways not previously mentioned.

4.2.3.1 Regulatory guidance

4.2.3.1.1 The WP

The use of the phrase "risk-based approach" by the WP has several connotations. For example, it a phrase used by reference to the notion of "*a scalable proportionate approach to compliance*",[328] with explicit acknowledgement of the fact that there can be "*different levels of accountability obligations depending on the risk posed by the processing in question".*[329]

As mentioned, WP218 equates the concept of risk explicitly with "*potential negative impact on the data subject's rights, freedoms and interests*" to be determined taking into account specific

[327] As Tene points out (2011, p.7), adopting an approach that restricts the scope of the term personal data based on the risk of identification *"confirms to the spirit of Recital 26 of the Directive"*.
[328] WP218, p.2.
[329] Ibid, p.3. See also the assertion by the WP in the same document (p.2) about the risk-based approach: "*[t]he Working Party recognizes that* ***some of the provisions in the proposed Regulation may pose a burden on some controllers which may be perceived as unbalanced and has therefore in earlier opinions already expressed the view that all obligations must be scalable to the controller and the processing operations concerned.*** *Compliance should never be a box-ticking exercise, but should really be about ensuring that personal data is sufficiently protected. How this is done, may differ per controller…. Data subjects should have the same level of protection, regardless of the size of the organisation or the amount of data it processes. Therefore the Working Party feels that all controllers must act in compliance with the law, though this can be done in a scalable manner"* (emphasis added). This statement is a direct quotation from an earlier document - Article 29 Working Party (Statement of 27 February 2013, pp.2-3).

objective criteria, including "*the nature of personal data*"[330] and "*the purpose of the processing*".[331] It also notes, "*[t]he severity and the likelihood of the impacts on rights and freedoms of the data subject constitute elements to take into consideration to evaluate the risks for individual's privacy*".[332]

WP136 contains similar risk statements specifically regarding the concept of personal data:

> The objective of the rules contained in the Directive is to protect the fundamental rights and freedoms of individuals, in particular their right to privacy, with regard to the processing of personal data. **These rules were therefore designed to apply to situations where the rights of individuals could be at risk and hence in need of protection**…This is a very important element to take into account in the interpretation and application of its rules. It may play a substantive role in determining how to apply the provisions of the Directive to a number of situations where the rights of individuals are not at risk, and it may caution against any interpretation of the same rules that would leave individuals deprived of protection of their rights.[333] (emphasis added)

Unfortunately, in both publications, the WP neither expands on the types of risks that are concerning ('risks of what happening precisely?') beyond the fact that they must have some impact on (i.e. affect) individuals' rights, nor does it give associated examples of when data processing may have negative impact on their rights.

Yet in WP203, the WP equates risk with consequences (i.e. effects).[334] Moreover, in that same document, the WP also links explicitly the notion of 'impact' and the concept of personal data, with

---

[330] As mentioned above in relation to sensitive personal data, from this quote it seems that, even where the WP affirms that a category of data should be deemed personal data, it acknowledges that the processing of different data types can provide different degrees of threat to an individual. See also its opinion in respect of processing activities applied to data within the same data category (Article 29 Working Party, 2011, WP185); the WP stated (at p.17) that – while it recommended that a data controller should treat "*all data about WiFi routers as personal data*" – the mapping of WiFi access points in principle (given their semi-static nature) constitutes a "*lesser threat to the privacy of the owners of these access points than the real-time tracking of the locations of smart mobile devices*".

[331] WP218, p. 4 (in full): "*[r]isks, which are related to potential negative impact on the data subject's rights, freedoms and interests) should be determined taking into consideration specific objective criteria such as the nature of personal data (e.g. sensitive or not), the category of data subject (e.g. minor or not), the number of data subjects affected, and the purpose of the processing … The severity and the likelihood of the impacts on rights and freedoms of the data subject constitute elements to take into consideration to evaluate the risks for individual's privacy*…"

[332] Ibid, p.4. On the other hand, the WP equates a risk-based approach in that context as involving much wider deliberations related to effects (p.4): *"[t]he risk-based approach goes beyond a narrow "harm-based-approach" that concentrates only on damage and should take into consideration every potential as well as actual adverse effect, assessed on a very wide scale ranging from an impact on the person concerned by the processing in question to a general societal impact (e.g. loss of social trust)".* Chapter 7 below revisits this point.

[333] WP136, p.25.

[334] Article 29 Working Party (2013, WP203, p.18): "*[a]n example of an unacceptable high residual risk includes where the data subjects may encounter significant, or even irreversible, consequences, which they may not overcome, and/or when it seems obvious that the risk will occur*." See also p.8, where – in listing "*automated-decision making with legal or similar significant effect*" as an example of high-risk processing – it gives the example of processing which "*may lead to*

its admission that "*processing of personal data has an impact on individuals' fundamental rights in terms of privacy and data protection. This impact on the rights of individuals must necessarily be accompanied by a limitation of the use that can be made of the data".*[335] Whereas, on the next page, the WP retreats from the language of rights to a more general statement that *"data protection law …has after all been designed to protect individuals against the impact **of** improper or excessive use of their personal data"* (emphasis added).[336]

Despite this inexact phraseology,[337] elsewhere the WP explores the concept of risk-impact in other guidance it has published – including pre-GDPR on the topic of privacy impact assessments (PIAs). (Section 4.2.3.2 below addresses the related concept of data protection impact assessments (DPIAs)). PIAs come in a variety of forms, but, fundamentally, they are all strategic planning exercises associated with assessing the likely privacy impact of future data processing activities.[338] Said otherwise, they proscribe the ex-ante assessment of likely negative effects associated with processing data relating to persons in specific instances (presumably, in many data protection scenarios after the decision has been taken about whether personal data would be involved, although as a matter of practicality this may not always be the case). Additionally, PIAs encourage data holders to think about ways in which to mitigate any adverse privacy effects subsequently identified. As Raab comments, the latter dimension of the PIA process – enabling the subsequent management of any likely negative privacy impact on individuals found - plays an important role.[339]

The WP has issued guidance incorporating advice around the legal basis for a PIA framework promulgating its use, in turn helping organisations to carry out reasonably objective assessments around likely privacy impacts from their data processing plans and manage these (especially regarding new technologies).[340]

---

the exclusion or discrimination against individuals", whereas *"processing with little or no effect on individuals does not match this specific criterion*".

[335] Ibid, p.24.

[336] Ibid, p.25. The WP says directly afterwards: *"[t]he nature of the data processed plays a critical role in all its provisions. It would therefore be important to evaluate whether the further processing involves sensitive data, either because they belong to the special categories of data under Article 8 of the Directive, or for other reasons, as in the case of biometric data, genetic information, communication data, location data, and other kinds of personal information requiring special protection".*

[337] Discussed further in Chapter 7 below in terms of a possible evolving theory of 'risks to rights' emerging under data protection law, as discussed in Van Dijk et al (2016).

[338] This exercise format allows those who intend to process data relating to persons to address any issues prior to implementing its activities (comparisons may be drawn here with notions of 'privacy by design' and 'privacy enhancing technologies') From a historical perspective, PIAs have been around since the late 1990s in a few countries, notably Australia, Canada, Hong Kong, New Zealand, and the US.

[339] Raab (2005, p.14): *"[a] further definition of PIA illustrates an important dimension: "[PIA] is an analysis of how information is handled: (i) to ensure handling conforms to applicable legal, regulatory, and policy requirements regarding privacy; (ii) to determine the risks and effects of collecting, maintaining and disseminating information in identifiable form in an electronic information system, and (iii) to examine and evaluate protections and alternative processes for handling information to mitigate potential privacy risks".*

[340] So has the European Commission in tandem. For example, it published an EC Recommendation in May 2009 regarding the use of PIAs in new RFID applications; the WP then subsequently endorsed industry-prepared guidance in

4.2.3.1.2 The ICO

The ICO also promotes PIAs as a useful compliance tool, arguing that organisations choosing to conduct them before engaging in new data processing activities are better prepared. The ICO has published numerous guidance with recommendations on how to conduct structured, legal compliance checks against DPA requirements using PIAs.[341] See, for example, the ICO's PIA Handbook,[342] plus a more recent Code of Practice on Conducting PIAs.[343]

In its Report on Big data and Data Protection, the ICO further mentioned PIA as a tool for compliance available to organisations so they can "*understand how the processing will affect the people concerned*",[344] "*to what extent it is likely to affect the individuals whose data is being used*",[345] and thereby ensure that "peo*ple's data privacy rights are respected*".[346] The ICO also deems it a corollary and necessary part of PIAs for assessors to analyse whether, after the application of mitigatory solutions, the final impact on individuals in respect to identified possible privacy impacts is proportionate to the aims of the planned processing operation.[347] A related 2015 ICO

---

respect of the PIA Framework in the field of RFIDs (WP, 2011) after a multi-stakeholders consultation and drafting process. The European Commission has also funded a privacy impact assessment framework (PIAF) project, completed in 2012, which offers guidance to organisations on how to best carry out a PIA and what benefits they can expect in doing so.

[341] Although, to note, the DPA does not require organisations to conduct PIAs (nor does the DPD). Notwithstanding, and while PIAs appear prima facie intended to assess the implications on the fundamental right to privacy, yet what they in fact evaluate typically are implications on the right to data protection, when dealing with the processing of data. The fact that PIAs focus on data protection issues rather than on various aspects of privacy has also been raised by other scholars. See, e.g. Milaj (2015), and De Hert et al (2012).

[342] ICO (2007, rev.2009).

[343] ICO (2014, #1). It gives practical advice on how to do PIAs (including reference to a broad coverage of impact-effect types - physical safety, material, and moral – as well as describing ways in which the likely privacy impacts of processing activities on individuals can be analysed), and links the structure of assessment to standard risk-management methodologies. The Code of Practice also emphasises the importance of documenting likely privacy impact risks, and determining strategies for their possible mitigation, as an important part of the overall PIA process.

[344] ICO (2014, #1, para.20, p.5)**.** Compare ibid, para.100, p.31: *"[a]ssessing privacy risk involves being clear at the outset about the benefits and aims of the big data project, **as well as the impact on individuals' privacy.** In many cases, the benefits in question are benefits to the organisation that is proposing to process the personal data, but it is important to factor in also benefits that may accrue to individuals or to society more broadly"* (emphasis added). See also e.g. para.47, p.14 - considering effects: "*[f]airness involves a wider assessment of whether the processing is within the reasonable expectations of the individuals concerned. For example, every aspect of analysing loyalty card data to improve marketing should not always automatically be considered fair, or within customer expectations".* To note, in the updated version of this publication (ICO, 2017, #2), there are also references to effects (at paras 31-32, pp.19-20): "*assessing fairness also involves **looking at the effects of the processing on individuals**, and their expectations as to how their data will be used... How big data is used is an important factor in assessing fairness. Big data analytics may use personal data purely for research purposes, eg to detect general trends and correlations, or it may use personal data to make decisions affecting individuals"* (emphasis added). See also, para 38, p.22*: "[t]his means that if big data organisations are using personal data, then as part of assessing fairness they need to be aware of and factor in the **effects of their processing** on the individuals, communities and societal groups concerned. Given the sometimes novel and unexpected ways in which data is used in the analytics, this may be less straightforward than in more conventional data-processing scenarios. Privacy impact assessments provide a structured approach to doing this"* (emphasis added).

[345] ICO (2014, #1, para.99 p.30).

[346] Ibid, para.98 p.30.

[347] Ibid, para.100, p.31: "*[w]hen solutions to mitigate privacy risk have been identified, it is necessary to assess whether the final impact on those individuals, after those solutions have been applied, is proportionate to the aims of the project".*

publication[348] also echoes the importance of assessing the impact of analytics on individuals, as well as "*differentiating between levels of impact*".[349]

Separately, but using similar terminology, in 2014 the ICO discussed the risk of likely impact flowing from information processing in relation to the boundaries of the concept of personal data. It links this to the threat of effects on individuals flowing from the possibility of their (re-)identification from information (see Chapter 5 specifically on the issue of data de-identification/re-identification):

> [T]he DPA does not apply to anonymised information – reflecting the low risk this form of information poses. The logic is that **once information ceases to identify anyone then it ceases to have any direct effect on them, so it poses a lower privacy risk**.[350] (emphasis added)

This statement is bold and arguably incorrect: information that identifies no-one can have huge effects on people (such as flowing from certain statistics used in certain contexts, see the cases in Appendix 3). Of course, the ICO may be placing weight on the notion of the risk being privacy-related here in explaining its logic for making this statement. To dig deeper into their rationale for this statement, it is useful to back-track to much earlier (2001) ICO interpretational guidance on personal data under the DPA. There, the ICO expressed the view that web information is personal data where used with "***an intention to target that particular user with advertising, or to offer discounts when they re-visit a particular web site,*** *on the basis of the profile built up, without any*

---

[348] IC (2015).

[349] Ibid, p.3. See also p. 9: "*[w]e agree that PIAs are particularly important in the context of big data analytics. We will continue to promote our Privacy impact assessment code of practice which contains practical advice on how to do PIAs. One respondent argued for the importance of privacy risk assessments: they can enable responsible decisions about data use, they place the burden of privacy protection on the organisation and they allow for flexibility in the application of the data protection principles. We agree with these points and we think that the principles of a privacy risk assessment, as described, are very much in line with those of PIAs. We will liaise with key stakeholders to discuss the development of more specific PIA guidance on big data that uses the ICO PIA code as a framework. We would look to identify a sector, professional or industry body to take this work forward. This should also be supplemented by case studies"*.
Furthermore, the ICO commented in its press release (no longer available online) accompanying the publication of the Summary of Feedback in response to ICO, 2014, #1: "*[a]nother theme which emerged strongly [from respondents' comments] was the importance of having a framework for assessing and mitigating privacy risks****".***

[350] ICO (2014, SMD0018). For further discussion and relevant ICO quotes, see Chapter 5. For example, see its Anonymisation Code of Practice (2012), at p.40 ("***As anonymised data has no direct effect on any individual…***"), and p.58 (***"****[processing personal data] will not breach the data protection principles as the purpose of the redaction process is to protect the individual research subjects' privacy and* ***the processing itself has no direct effect on any individual****")* (emphases added). This ICO Code further highlights the importance of secondary factors in informing an organisation's approach to data disclosure and security, such as the type of data at issue and who it is about (both factors of which could impact upon the implications for an individual in case they are re-identified), p.23: "*[c]learly, some sorts of data will be more attractive to a 'motivated intruder' than others. Obvious sources of attraction to an intruder might include: for nefarious personal reasons or financial gain;* ***the possibility of causing mischief by embarrassing others;*** *revealing newsworthy information about public figures; political or activistic purposes, eg* ***as part of a campaign against a particular organisation or person****; or curiosity, eg a local person's desire to find out who has been involved in an incident shown on a crime map"* (emphasis added).

ability to locate that user in the physical world" (emphasis added).[351] By comparison, in that same year, the then Information Commissioner, Elizabeth France, commented that: "*[i]f static IP addresses were to form the basis for profiles* ***that are used to deliver targeted marketing messages to particular individuals*** *they, and the profiles, would be personal data subject to the [DPA]*" (emphasis added).[352] Said otherwise, as far back as the early 2000s, high-level voices within the ICO acknowledged an underlying concern related to the processing of data relating to persons going beyond an ability to distinguish someone from that data per se, towards a recognition of impact-effects as being the real underlying threat justifying application of the regulatory regime.

4.2.3.2 The GDPR

The GDPR too promotes risk-based assessments as a means for organisations handling personal data to demonstrate accountability (as described in its Article 5(2)),[353] alongside the need to implement ex-ante risk-preventative measures such as the principle of 'data protection by design and default' (Article 25). For example, Recital 76 GDPR states, *"[t]he likelihood and severity of the risk to the rights and freedoms of the data subject should be determined by reference to the nature, scope, context and purposes of the processing. Risk should be evaluated on the basis of an objective assessment, by which it is established whether data processing operations involve a risk or a high risk*."[354] Again, however, it is not clear from a risk-management perspective to this statement how

---

[351] ICO (2001, p.12) (in full): "*[i]f the information about a particular web user is built up over a period of time, perhaps through the use of tracking technology, with the intention that it may later be linked to a name and address, that information is personal data. Information may be compiled about a particular web user, but there might not be any intention of linking it to a name and address or e-mail address.* ***There might merely be an intention to target that particular user with advertising, or to offer discounts when they re-visit a particular web site,*** *on the basis of the profile built up, without any ability to locate that user in the physical world. The Commissioner takes the view that such information is, nevertheless, personal data.* ***In the context of the on-line world the information that identifies an individual is that which uniquely locates him in that world, by distinguishing him from others****"* (emphasis added). Notwithstanding, as mentioned, it is possible that the ICO mixes up discussions of the 'relating to' element and the 'identificatory' requirement. For example, compare the TGN, p.17: "*data may be personal data because it is clearly 'linked to' an individual because it is about his activities and is processed* ***with the purpose of*** *determining or influencing the way in which that person is treated*" (emphasis added).

[352] Quoted from Pinsent Masons (2008, IP addresses and the Data Protection Act? [online]).

[353] Article 5(2) defines 'accountability' in terms of controllers being "*responsible for, and be[ing] able to demonstrate compliance with, Article 5(1)*" [setting out the general data protection principles mirroring those found in the DPD: *'lawfulness, fairness and transparency'; 'purpose limitation'; 'data minimisation'; 'accuracy'; 'storage limitation'; and, 'integrity and confidentiality*']. In effect, it requires proactive and demonstrative compliance with data protection obligations in practice, including by way of adoption of internal policies and mechanisms to help achieve such compliance. Arguably, such measures as taken should be those most appropriate to the specific circumstances - in particular in light of the risks posed by the data processing intended - considered right from the start of the planning process. Compare comments in an opinion on the principle of accountability by the Article 29 Working Party (2010, WP173, p.4-5) regarding the need for the provision of legal certainty in data protection law, while also aiming for rules "*formulated in sufficiently broad terms to allow scalability (enabling the determination of the concrete measures and verification methods to be applied depending on risk of the processing and the types of data processed))… the type of procedures and mechanisms would vary according to the risks represented by the processing and the nature of the data".*

[354] Compare Recital 77, GDPR: "*[g]uidance on the implementation of appropriate measures and on the demonstration of compliance by the controller or the processor,* ***especially as regards the identification of the risk related to the processing, their assessment in terms of origin, nature, likelihood and severity, and the identification of best practices to mitigate the risk, could be provided in particular by means of approved codes of conduct, approved certifications,***

processing "risk" in that context is meant to be particularised (i.e. risk of what exactly and in what ways, beyond 'risk to the rights and freedoms of individuals'?).[355] In other words, the range of risks and freedoms of individual (e.g. under human rights law, if this is what is being referred to here) is so large that arguably any action could potentially trigger their relevance and potential conflicting relevance (e.g. right to privacy versus a right to free speech). Arguably, the risk should be of something happening, e.g. 'a risk of interference with…likely to cause such and such end', or – better – a risk of harm to individuals felt concretely that would be verifiable.

More pertinently, however, the GDPR specifically mandates impact-based risk-assessments to be undertaken by data controllers, in turn associating such risks as potentially leading "*to physical, material or non-material damage*" (Recital 75). Like PIAs,[356] but now rebranded data protection impact assessments (DPIAs), a DPIA is a methodological exercise for risk-assessing upfront the likely impact of data processing activities for use where they are likely to raise a "*high risk to the rights and freedoms of natural persons*" (Article 35 GDPR).[357] Thus, from 25 May 2018, extensive use of contextual effects-based type data processing assessments will be mainstream as it is elevated from being an existing area of good practice (a recommended compliance tool), to a legal obligation in certain risk assessed circumstances.

Article 35 GDPR supplies non-exhaustive examples of high-risk categories of data processing activities.[358] As considered more fully in 2017 WP draft guidance on how to carry out DPIAs

---

guidelines provided by the Board or indications provided by a data protection officer. *The Board may also issue guidelines on processing operations that are considered to be unlikely to result in a high risk to the rights and freedoms of natural persons and indicate what measures may be sufficient in such cases to address such risk*" (emphasis added).

[355] A similar reference to risk "to the rights and freedoms of natural persons" is mentioned in Article 33 GDPR (concerning notification of a personal data breach to the supervisory authorities): "*1.In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent in accordance with Article 55,* ***unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons"*** (emphasis added).

[356] With which continuity is envisaged, see Article 35(8) GDPR: "*[c]ompliance with [DPA] approved codes of conduct referred to in Article 40 by the relevant controllers or processors shall be taken into due account in assessing the impact of the processing operations performed by such controllers or processors, in particular for the purposes of a data protection impact assessment."*

[357] Article 35 GDPR states: "*[w]here a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data*". It also enumerates a prescriptive list of what a DPIA should contain—a general description of the envisaged processing operations, an assessment of the risks, the measures envisaged to address the risks, demonstrated compliance with the Regulation, and a requirement to seek the views of data subjects. Where a DPIA indicates that there is a high degree of specific risk, the DPA's authorisation must be obtained under Article 34 GDPR: "*[a] controller/processor may also voluntarily request the prior authorisation of the DPA for its proposed processing in order to verify compliance".*

[358] Article 35(3): *"(a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person; (b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or (c) a systematic monitoring of a publicly accessible area on a large scale*". See also Article 35(4):"*[t]he supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment pursuant to paragraph*

(WP248)[359] discussed in the next section, the main elements of assessment required to analyse the impact risk posed by particular processing are set out in Article 35(7):

> (a) a systematic description of the envisaged processing operations and the purposes of the processing… (b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes; (c) an assessment of the risks to the rights and freedoms of data subjects …; and (d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.

Therefore, in respect of risks thought likely to have a negative impact on data subject, Article 35 explicitly mentions the taking of remedial actions as necessary to minimise such effects. It also contains some ancillary obligations, such as consulting the supervisory authority in certain circumstances, prior to processing.[360] Moreover, the GDPR conceives this tool as something to be used on an on-going basis, including reviews of ex-post activity deployment.[361]

4.2.3.3 Summary

There is evidence of the acceptance of effects-based concepts already in data protection law/regulation along with encouragement of risk-based assessments generally. Specifically, the PIA and the new DPIA are akin to effects-based assessment tools for implementation before processing, in common with the Effects-based Approach. Moreover, even though PIAs/DPIAs would often take place after the issue of whether data are personal are considered already settled, insights from their underlying methodology may still be considered useful as part of the definitional (i.e.

---

1. The supervisory authority shall communicate those lists to the Board referred to in Article 68. The supervisory authority may also establish and make public a list of the kind of processing operations for which no data protection impact assessment is required. The supervisory authority shall communicate those lists to the Board" (emphasis added)

[359] Article 29 Working Party (2017, WP248). An analysis of that document is carried out by this author in Knight (2017, New EU guidelines on data protection impact assessments, [online]). The WP adopted a revised version of that document on 4 October 2017 (2017, WP248 rev.01). Like the draft version, the latter document also emphasises how controllers must continuously assess the risks created by their processing activities. This obligation is to ensure that they can identify when a type of processing is likely to result in a high risk to individuals, but also – even where conditions where conditions triggering the obligation to carry out a DPIA are not met - data controllers should regard themselves as under "*a general obligation to implement measures to appropriately manage risks*" for individuals (see p.6).

[360] Article 36 GDPR provides: "*[t]he controller shall consult the supervisory authority prior to processing where a data protection impact assessment under Article 35* ***indicates that the processing would result in a*** *high risk* ***in the absence of measures taken by the controller to mitigate the risk*****"** (emphasis added). Additional measures required when specific risks are identified (e.g. impact assessment, enhanced security, data breach notification) are provided for in Articles 33 and 34 GDPR.

[361] Article 35(11) GDPR: *"[w]here necessary, the controller shall carry out a review to assess if processing is performed in accordance with the data protection impact assessment at least when there is a change of the risk represented by processing operations".*

jurisdictional – 'are my planned processing activities subject to data protection law first of all, in any event?') determinations that any data holder must carry out pre-processing.[362]

4.3 Critical assessment of the Effects-based Approach

Justification for the Effects-based Approach may draw inspiration, not only from an analysis of the pertinent policy guidance as it develops within the evolving data protection legislative framework, but also from the interpretative work of legal scholars. This includes literature that advocates models of data protection law incorporating effects-based considerations, which will be compared to but also distinguished from the main thesis ideas.

4.3.1 Legal literature review and how legal scholarly theories situate alongside a legal proposal for an effects-based approach

This sub-section sets out a critical discussion of a selection of legal literature in this area, including by the main legal scholars who are proponents of effects-based theories – or who propose theoretical variants on an effects-based theme in their legal models. While illustrating that the main ideas in this thesis are feasible, this review will also endeavour to show that such ideas are demonstrable better in practice than those proposed by others. Moreover, this analysis will help demonstrate a systematic acquisition and understanding of a substantial body of knowledge which is at the forefront of the legal academic sub-discipline, and area of professional practice, that is data protection.

Booth et al

In reminder, the Effects-based Approach is suggested in this chapter as an alternative to the current data protection regime's reliance on identification and identifiability as keystones of the definition of personal data. A similar type of proposal was deliberated by Booth et al (2004) who conducted a study for the ICO analysing the extent to which - based on varying interpretation of the DPD – there had emerged different concept of what constitute personal data across the EU MSs. In other

[362] In other words, when determining whether data relating to persons are personal data pursuant to the Effects-based Approach, the terminology and elements of the methodology of traditional PIA and new DPIA risk-assessments may be helpful broadly in assessing the potential harms to data subjects likely to result from the processing of such. However, carrying out a comprehensive PIA/DPIA every time that data relating to persons is being considered would go beyond the scope of this recommendation. Said otherwise, it is being proposed that a Relevant Effect risk analysis should be considered by those intending to process data relating to persons to assess the potential impact of each intended data processing activity on such persons. However, this is not intended to amount to a requirement that a full PIA/DPIA be carried out in each case but rather a suggestion that insights could be gained from its methodology in adaptation for its own methodology development, in particular when this would be useful in relation to 'grey area' cases that warrant it from a risk perspective. The key element of the Relevant Effect analysis framework for development are discussed next in Section 4.3.

words, the study concluded that there was no coherent definition of personal data applied consistently pan-EU (at least with respect to the period up to the early 2000s). The authors found that three different personal data legal conceptual approaches – what they call 'ideal types' or 'classificatory models' – had emerged:

1) Data-identificatory potential as a prerequisite;

2) Data-effects potential as a prerequisite; and,

3) Identification and effects as prerequisites (a composite approach requiring both features to be present in a data processing situation).

However, their analysis – and, in particular, their second effects-based model of personal data (i.e. that data must be an effect on a person's privacy for it to be construed personal data legally) - was carried out in respect of interpreting the 'relating to' element of the legal definition of personal data under the DPD. Their research premise is therefore different from the current thesis study.

Notwithstanding, relevant aspects of this study-report bear similarities to those considered in this thesis. For example, Booth et al – like this thesis author – highlight that:

- Context is all important, whereas a classificatory model based around effect may need to assume relatively stable contexts taking note of the various uses to which the data could be put and the impact of such uses – i.e. taking into account how a piece of data could ordinarily be used within a typical informational context (p.16).

- *"The development of any robust theoretical framework capable of justifying and supporting decisions about the boundaries and definition of personal data' cannot take place within a vacuum*" but must "r*emain grounded within the realities of practice*" (p.100).

The latter aspect, in particular, is highlighted in this thesis as any conceptually coherent legal definition of personal data should be, not just justifiable but also practical. Moreover, Booth et al conclude that distinguishing an identificatory and an effects-based ideal type to construing (and making decisions in practice) whether information falls under data protection laws on processing may provide an incomplete picture (an incomprehensive understanding) without more. To overcome the perceived limitations of each model, the authors appear to prefer a composite concept of personal data in practice for usage in underpinning any decision making strategy despite it being "*bound to be a challenging task*" (p.128). This would entail isolating both a particular way that privacy would have to be affected, and a particular circumstance within which identification would have to take place.

At the end of this chapter, too, the idea is put forward that the disadvantages felt if relying upon any one of these ideal types exclusively to make decisions on personal data in practice suggests a better strategy is one that extracts the best bits of both approaches – acknowledging the deep ingraining of the identificatory approach in legal discourse and in the GDPR text - when developing this thesis' unique contribution to knowledge in later chapters.

Ohm

Ohm's well-known 2010 paper on anonymisation was influential in setting a discourse running about the failures of reliance upon the concept of anonymisation as a guarantee of privacy. This issue will be discussed at length in the next chapter.

For now, by way of a summary of his main ideas and as mentioned earlier in Chapter 1, Ohm presents an (re-)identificatory approach in conceptualising the concept of (non-)personal data in the context of the public policy debate about information privacy. Precisely because of the prevalence of (re-)identification risks involved when processing any data relating to persons in this 'big data' era, Ohm concludes that the best course is to remove completely the personal data (/PII) definition in privacy laws.

This proposition to do away with a jurisdictional gateway concept is problematic and at odds with the Effects-based Approach as argued in this thesis. In effect, if applied to EU data protection law, Ohm's proposal would mean that the law would be left without a 'flag' for establishing a coherent boundary to necessary regulation. In turn, this would deprive those handling data relating to people of any sense of legal certainty as to whether they need to comply with a body of law or not.

Notwithstanding, Ohm does make some suggestions for alternative regulatory restrictions that could apply, some of which are similar to certain arguments set out in this thesis. Broadly, Ohm argues that those enacting privacy laws should consider the benefits versus the costs of information usage, and suggests that they be minded to only impose restrictions when such costs outweigh the benefits of an unfettered information flow to societal values in context. More specifically, Ohm proposes a new regulatory scheme for deciding whether to impose privacy restrictions underpinned by a test assessing the likely consequences resultant from prospective information usage in context.

Ohm's test would involve analysis of five factors to assess the potential risks arising in a particular data-processing context in relation to the following issues: 1) 're-identification risk' (the likely impact of the data-handling techniques to be used on the risk of re-identification from the relevant information); 2) 'public or private' (whether the information is being released publicly or privately); 3) 'quantity' (the amount of information to be released); 4) 'motive' (motivations for releasing the

information); and, 5) 'trust' (the amount of trust to be placed in those in current possession of the information). From considering these five factors cumulatively, Ohm argues (2010, p.1768) that regulators should be well-placed to exercise their judgement and intervene only when they assess that the risks from the prospective information usage by a particular type of data handler in a particular context are high. On the other hand, Ohm says regulators may choose to do nothing if the risks are assessed to be low such that they do not outweigh the benefits from a potential unfettered information flow.[363]

In common with the main ideas in this thesis, Ohm's theory endorses carrying out a risk-assessment as a regulatory yardstick for determining the application of data protection obligations. However, he couches such discussions in generalised language that omits reference to effects and, while mentioning harm, he does so often in an oblique manner (e.g. suggesting regulation should apply if "*the benefits of unfettered information significantly outweigh the costs to privacy in a particular context*", p.1736), and is also sketchy on detail as to the implementation of his approach (e.g. *"[o]nce the rules of verifiable trust are codified, regulators can free up data sharing between trusted parties*", p.1770). Furthermore, the risk-assessment multi-factorial test Ohm sets primarily list re-identification risk – thus retaining an identificatory approach narrative – while, for example, the factor of trust to be placed in the data possessor he suggests is highly subjective and thus unsuitable for use in a regulatory environment.[364] By contrast, this thesis focusses primarily on the negative (and positive) effects likely to arise specific to the processing-usage activity under consideration, for assessment ex ante.

While Ohm acknowledges the profound social benefits that can result from data practices and flows, so does the Effects-based Approach in recognising that data processing can lead to positive effects (including effects on collective interests) that may also be considered in determining the outer limits of the application of data protection rules. However, in unravelling this acknowledgement, Ohm appears to relegate privacy law fundamentally to a direct cost-benefit trade off that is strictly utilitarian (where harm from processing could potentially be appreciable and yet still permitted because of the extent of the benefits likely from unfettered information

[363] To demonstrate how a regulator would apply this test, Ohm discusses two case studies involving health information, and internet usage information, respectively. He uses these to illustrate how regulation might adapt to using his new risk-assessment strategy in each context discussed bearing in mind the specific types of benefits that might ensue under each case study.

[364] For example, Ohm suggests at one point that milder restrictions could be placed on small classes of trusted people while banning the sharing of the data with anybody else. Arguably, however, it is the usage of data – not the type of people/organisations involved – that matters when it comes to assessing harm impact possibilities in context (as one class of people/organisation can be minded to process data relating to persons for a whole range of different purposes).

flows), and liable to be skewed in practice (depending on the one doing the assessments and their self-interests).[365]

In development of the Effects-Based Approach put forward in this thesis, by contrast, the focus is on reducing the likely harm arising from specific data processing foremost, sitting alongside an acknowledgement of benefits potentially arising from specific processing activities *after* such harm mitigation measures are put in place. It does this via discussion of how to evolve the Effects-based Approach into a useable legal framework – and structured approach – in Chapters 5 and 6 that would allow for (re-)use of data by incentivising utility-maximisation, information sharing, and beneficial uses of analytics, all in ways that would be unlikely to raise appreciable risks of privacy harms. The formal framework and methodology for assessing effects according to objective standards, and confirmation of what criteria must be satisfied for data processing activities to be automatically legitimised, would be made by regulators (the DPAs) upfront.

Schwartz/Solove

Staying within the conceptual ambit of the identificatory-approach, US scholars Schwartz and Solove (2011, 2013) discuss and criticise at length the PII concept under US law. In particular, they propose a (re-)identification risk-based categorisation scheme, entitled "PII 2.0", to replace the binary 'PII/non-PII' distinction (across US, as well as EU law).

Inherent to this proposed scheme is a distinction between 'nominally identifiable information' ("in which linkage to a specific person has not yet been made, but where such a connection is likely" (2013, #1, p.907), and other types of identifiable information. Schwartz/Solove argue that, while nominally-identifiable data should be treated equivalently to identified data and subject to the full gamut of privacy and data protection rules, this is not true of other types of identifiable information from which specific identification of an individual, while possible, is not a significantly probable event. The latter should fall within an intermediary category of data in the middle of their risk continuum, which they propose should be subject to some but not all privacy restrictions. This is because the authors perceive the risk level associated with intermediate category data as low to moderate, which justifies information of this type being regulated differently. Moreover, they acknowledge that a data controller may put measures in place around nominally-identifiable data

[365] To his credit, Ohm recognises the tendency to skew risk/benefit predictions in practice. See, e.g., Ohm (2012, p.339) critiquing Google Flu Trends' limited real world impact and the ease with which supposed benefits can be cited to avert discussion of risk and the need for careful analysis of privacy risks and transparent engagement with users.

that lead to its classification changing so that it fell subsequently within the intermediate category.[366]

In other words, these US authors recommend within their analytical framework treating information with varying degrees of identifiability differently under the law; with different sets of obligations for different kinds of personal data. There are similar acknowledgement of the attractiveness of preserving a broad interpretation of personal data but reducing the intensity of compliance obligations commensurate with the degree of risk of harm in a particular processing context in Chapter 6 below. This notion supports the ultimate conclusion that the best direction of travel for EU data protection law is one where its full requirements need not apply to all types of processing involving data relating to people. Schwartz/Solove's more detailed discussions on this point are considered in Chapter 6 albeit also distinguished from this thesis author's ideas.

Schwartz/Solove also share other similarities with the ideas in this thesis, not least because of the challenge related to data 'accretion' problems in determining the point at which privacy/data protection should apply, as both acknowledge that periodic assessments of risk are needed to meet this challenge. Notwithstanding, as mentioned, Schwartz/Solove focus on the risk of re-identification and occasionally broadly conceived "privacy threats/risks" with little unpacking of this term, whereas this thesis focuses on the risk of negative effects resulting from data processing (while recognising the potential for positive effects simultaneously being generated). Notwithstanding, even Solove/Schwartz acknowledge under PII 2.0 that re-identification risk should be considered in the context of the particular analytics activity that is planned to take place, with PII assessable on a risk matrix taking into account the risk, intent, and *potential consequences* of re-identification.

Moreover, the discussions of Solove/Schwartz highlight how in practice, it is unlikely to be possible to entirely separate a narrative of (de-)identification from one of consequential harm (and harm mitigation) as in many cases the two narratives are often confused. For example, they state (2013, #1, p.915, emphasis added): "*when a* ***breach involves only identifiable data, the harm that the information can cause to individuals is much*** *less likely to occur. Harm can only occur when the party who obtains the data also knows how to identify it*". It is this confusion that provides reason

[366] Schwartz and Solove (ibid) give the following example: "*[a]n example of identifiable information under the PII 2.0 model would be the key-coded medical data that the Working Party discussed in its "Opinion on the Concept of Personal Data". Some or all of this information might never be identified. Depending on the risk scenario, there may be only a remote chance of future linkage to a specific person. As a further example, Kuner's discussion of [an example used by Kuner of a singled out but nameless individual] the Verdi-loving physician may represent merely identifiable information under PII 2.0. Kuner's hypothetical leaves much open regarding the "data controller." We know only that the data controller himself cannot identify the person to whom the information relates. If the data controller also has strong measures in place to protect the data from exposure to others, the PII 2.0 model would classify the information as identifiable, but not identified".*

to push broadly to change the legal narrative to effects and emphasise assessing processing risks and lowering these where possible, whereas currently the prevailing narrative about personal data is often about encouraging those who hold data relating to persons to hold it in the least identifiable form possible.

Separate from Schwartz, Solove has also put together a 'privacy taxonomy' (Solove, 2008) in order to assist the legal system in grappling with the concept of privacy. He believes that since the goal of the law is to have protections that adequately prevent and redress particular problems or risks, we need to first understand the problems in order to evaluate the effectiveness of the protections. In devising a taxonomy, he focuses on the activities that invade privacy and create problems comprised of four basic groups of harmful activities: information collection, information processing, information dissemination, and invasion. It is interesting to note that three out of the four groups relate to data handling activities, which illustrates that this is a good starting point when attempting to determine the type of harm that may take place.

Mantelero

As mentioned at the start of this chapter, the term 'harm-based approach' has also been used in the context of describing a school of thought contending that the focus of privacy and data protection regulatory activity should centre on data usage. In that context, several legal scholars focus on the likely harms associated with processing activities, and allocation of legal responsibility where such harms have been facilitated.

Mantelero (2015) is one such legal scholar advocating a focus on risk-assessment as a model of data protection, interpreting this as requiring "*an assessment of the impact of…strategies on data protection…to reduce the potential negative effects on individuals*".[367]

Cate/Mayer-Schönberger

These co-authors also call in a paper (2013) for reducing "*the focus on data collection and the attending notice and consent requirements, and focus(ing) more on a practical assessment of the risks (and benefits) associated with data uses*".[368] Indeed, they reiterate elsewhere that the assessment of effects of processing should encompass both negative and positive potential impacts in tandem, because it is "*[o]nly by understanding and applying the tools of risk assessment and*

[367] Mantelero (2015, p.309): "*[i]n terms of data protection, this means the adoption of a new paradigm, which focusses on risk assessments…In the proposed model, companies that intend to adopt strategies based on Big Data with regard to personal information should conduct an assessment of the impact of these strategies on data protection, in order to adopt all the adequate measures and standards to reduce the potential negative effects on individuals and the risks of social surveillance"*. Compare, Mantelero (2014, pp. 1-7).
[368] Cate & Mayer-Schönberger (2013, #1, p11).

management to clearly articulated categories of potential harms and benefits likely to result from uses of personal data can we ensure that data are used responsibly, that individual privacy is protected, and that data users are accountable stewards of the data they possess".[369]

This argument is shared by this thesis author. However, Cate/Mayer-Schönberger – like Mantelero – do not 'fit' this line of thinking specifically in the context of discussing the legal definition of personal data (in trying to delineate what personal data should encompass and what it should not).

McCullagh

Like the majority of other EU legal scholars writing in this area, McCullagh (2009) tackles the question of when data are 'personal data', legally, foremost using an identificatory risk-based approach. Notwithstanding, McCullagh also mentions a potential approach that – like the Effects-based Approach – would involve consideration of likely harm that may be generated from the processing of data relating to people in context (e.g. taking into factors such as the data holder's purpose in carrying out a processing activity) in determining data's legal status under data protection law.

McCullagh's proposal is limited, however, to re-determining the legal sub-concept of sensitive personal data under the DPD as discussed later in this chapter, rather than being about the concept of personal data (and jurisdictional boundaries) per se. Nevertheless, she provides some useful 'food for thought' in prelude to the Effects-Based Approach. In particular, McCullagh's argument that retaining a legal category of sensitive personal data (re-named 'special category data' under the GDPR, Article 9) may become practically ineffective in the light of technological advances and societal changes presages the need for the reconceptualisation of the outer limits of the concept of personal data in ways that can work with future trends in terms of flexibility and usability (or else it would become redundant).

Tene

Tene (2011) suggests restricting the scope of the term personal data based on the likelihood of identification with a context-specific test for assessing risk. Tene is critical of an overly narrow definition of personal data that could overlook the potential for increasingly sophisticated means of singling out individuals via processing their data. To this end, he suggests that data protection's outer boundaries should be viewed as a fluid continuum with varying obligations depending on how far your processing sits along the scale, as opposed to the current dichotomy.

[369] Cate & Mayer-Schönberger (2013, #2).

Notwithstanding, Tene's ideas remain within an identificatory-approach analysis – such that he suggests that data which are only identifiable at great cost would be subject to only a subset of data protection principles, albeit they would remain within the scope of the law. That is, the degree of protection provided by data protection law should be scoped proportionately to the degree of data (re-)identifiability. On the other hand, Tene, in conjunction with Polentsky (Polentsky/Tene, 2013, p.258) also argues that, "*a bi-polar approach based on labelling information either personally identifiable or not*" is inefficient.

This comment, like the ideas in this thesis, suggest that the focus of the legal narrative should be on highlighting the essential issues underpinning the tussle of words and arguments by re-identifiers and de-identifiers in labelling information as personal or not to suit their ends (see Chapter 5 below). In other words, how can law/regulation help promote societal benefits flowing from certain processing-usages, while also ensuring that data utility does not come at the cost of other informational values (e.g. quality/accuracy). This is a challenge that an effects-based block exemption/safe harbour model – see in particular Model 2 described in Chapter 6 below - attempts to meet head-on, in promoting an outcome of maximum utility and benefits from data processing relating to people without causing appreciable harm.

Gratton

Gratton, a Canadian legal scholar, wrote a paper (2013) entitled, "If Personal Information is Privacy's Gatekeeper, Then Risk of Harm Is the Key: A Proposed Method for Determining What Counts as Personal Information". Broadly, this paper argues that data should fall under the definition of 'personal information' under Canadian law (which is broadly similar to the EU personal data concept) if it is liable to trigger a harm for the relevant individual irrespective of identifiability. In other words, there are striking similarities between the main ideas of Gratton's paper and this thesis.

Ultimately, Gratton suggests retaining the identificatory definitions set out in privacy/data protection laws to trigger their application, but construing them primarily using a risk of harm criteria. Said otherwise, Gratton suggests that the practical exercise of determining whether data identifies or is identifiable of a person should be swapped by an assessment test of the risk of harm (i.e. an effects-based approach) in the processing context. However, this swapping would be done under the primary 'banner' of the existing identificatory-approach, rather than heralding explicitly a new-approach decision-making strategy (which would aid transparency, accountability and predictability).

Gratton's key contribution is in proposing a framework with criteria to be considered indicating when data may create a risk of harm for individuals. This 'decision tree' format is intended to help

organisations decide when information should be held to qualify as personal data. She also suggests that this approach may pave the way for a more flexible framework than currently exists, which is necessary to address adequately the different types of harm that result from processing information related to people. In other words, for Gratton, interpretations of identificatory concepts in data protection/privacy laws should be moderated such that associated processing obligations should be less stringently applied where data presents a lower risk of harm. This places emphasis on the need to judge the seriousness of effects predicted.

In comparing these key aspects of Gratton's work with the main ideas in this thesis, both authors intend their approach to be used practically by organisations that handle data. Gratton too believes that assessing a risk of harm triggered by the handling of personal information is a function of several contextual variables that need to be considered in an integrated manner, such as the intentions of the parties involved in processing, the nature of the information (and the potential incentives for discovery and use by third parties within a given context), and the processing activities under contemplation. In terms of the metric of assessment (for whether data identifies an individual), Gratton draws a key distinction between data processing that may create an objective harm (being external) to an individual, and that which may create a subjective harm to them.[370] Similarly, this thesis author agrees that an effects-based risk assessment carried out ex ante should also encompass consideration of the potential for negative effects of an objective and/or sensitive nature following different types of processing activities.

For Gratton, objective and subjective harm typically arise in association with different types of processing activities as follows:

- **Objective harm** – Gratton identifies the '**usage**' stage of the risk-of-harm analysis as properly focused on whether objective harm (e.g., a financial harm, physical harm or some type of discrimination) might emerge out of a particular information use. In other words, it may be used to determine or influence the way in which a personal is treated or evaluated. She proposes (p.208): *"[i]f there is a negative impact (or what I refer to as an objective harm…), then the information would qualify as personal information and it would have to be "accurate" and "relevant" for the intended use". "*

[370] A distinction between two categories of privacy harm (one objective and the other subjective) is also made by Calo in his article (2011), The Boundaries of Privacy Harm. For Calo, the notion of objective privacy harm can be equated with a material, adverse consequence, such as when someone uses information about you against you to charge you a higher price, or deny you a service. By contrast, subjective harm (the knowledge of certain intimate details pertaining to an individual by another experienced as an injury) is akin to when you anticipate that someone will do something harmful or offensive to you in person, or the perception of being observed in a way that makes us feel threatened or uncomfortable.

- **Subjective harm** - Gratton (pp.177-178) describes this risk-of-harm analysis at the '**collection-storage/disclosure'** stage as assessing the likelihood of subjective harm arising. In other words, information may be construed personal data at this stage taking into account the potential for harm flowing to a person, involving a psychological type of negative effect on them (e.g. humiliation or embarrassment), if it were known by other persons. More specifically, Gratton proposes additional criteria relating to the information depending on the extent to which it is intrinsically intimate in nature, and the extent to which it is already available to third parties or in the public domain, along with evaluating the ease with which data correlation can occur.

To note, Gratton does not portray the two different 'stages' as hard and fast distinctions. For example, she suggests that the potential for harm linked to data-combining activities (a usage activity) depends on the nature of the information for combination (e.g. *"an IP address coupled with biographic information becomes more "sensitive", especially when linked with biographic information that is of an "intimate" nature and becomes more "sensitive" if this information was not already "available"…*"). However, to distinguish this thesis author's views, Gratton suggests that it is the act of combining an IP address with information of an intimate nature that makes it worthy of being deemed personal data (precisely because the disclosure of this information, if linked to an individual, would trigger a higher risk of subjective harm). Whereas, under the Effects-based Approach put forward in this chapter (and, as discussed in the next section, if the additional information is sensitive personal data whereby a legal presumption would arise that personal data is involved), it would be the prospect of the *activity of disclosure* itself (e.g. to an external third party if security around data storage is non-adequate) that may transform the *combined data* into personal data in anticipation of that imminent possibility, and not *merely* the act of combination.

Gratton also does not explore the degree of likelihood/severity of risk of harm that may be required for information to be deemed personal data, an issue that is explored by this author in the next section. Nor does Gratton consider the possibility of considering collective harm (also discussed in the next section) as part of her proposal, deeming the underlying issue outside the scope of data protection/privacy laws.

Partly, these omissions are due to the fact that Gratton is tethered to reinterpreting existing identificatory concepts that are grounded in the concept of individual uniqueness. This makes it difficult for her to break the bonds of the current narrative – such as to recognise that inevitably identifiability is doomed to be interpreted very broadly and should be accepted as such.

Whereas in this thesis, this weakness is recognised, as well as weaknesses associated with the Effects-based Approach related to lack of certainty and incentives for compliance. For this reason,

in Chapters 5/6, the main ideas of this chapter are elevated – from redrafting the current definition of personal data – to the proposition of a new regulatory tool (block exemption/safe harbour) based on an effects-based methodology as the preferred model outcome to address these shortcomings.

Hon et al

Some authors find an effects-based perspective appealing, particularly in approaching the concept of anonymisation. For example, co-authors Hon et al (2011) incorporate discussion of likely harm into their model of non-personal data. However, they do so primarily within the parameters primarily of a (risk-based) identificatory-approach to that concept, rather than under the primary 'banner' of a new approach.

They conclude that the personal data definition is no longer sustainable and present an alternative model, which is discussed in more detail in Chapter 5 below onwards. In brief, they advocate an approach to the definitional function of (non-)personal data that incorporates a harm-based perspective (i.e. considering potential effects). They argue that – while the definition of personal data should be based on the realistic risk of identification – "*the applicability of data protection rules should be based on the risk of harm and its likely severity*" from the particular processing (p.225). Thus, it should be considered in each particular context whether data protection rules should be applied and, then, which data protection obligations arise (the extent of which should be commensurate with the risk of harm likely arising, with more precautionary measures chosen as appropriate in the processing circumstances to address such risks). In this sense, Hon et al highlight how, pragmatically, forcing a direct choice between an identificatory approach and an effects-one may be inappropriate outside of a purely theoretical discussion.

In evolving the ideas of Hon et al, to make the effects-based part of their model more attractive, this thesis suggests a block exemption/safe harbour effects-centric tool that can be used to help those who would otherwise be left to choose. This thesis' preferred main contribution (Chapter 6, Model 2) proposes that a modulated approach to data protection obligations based on the level of harm likely from certain types of processing should be set out in such a legal instrument via a methodology to be followed. To distinguish Hon et al's model, the latter propose that consideration of which rules should be (dis-)applied should be made each context, whereas this would be pre-set within the terms of the block exemption regulation under Model 2.

Balboni et al[371]

[371] Balboni et al (2013).

Finally, Balboni et al (2013) argue that data protection legislation must find an "*appropriate balance*" between seeking to reconcile data subjects' rights and safeguards, and the free processing/flow of information across the EU, in line with societal expectations. They say this requires identifying ways to ensure processing is proportionate between the protection of individuals' fundamental rights and the demands of the general interest of the community. To this end, they propose that lawful processing should be subject to an explicit 'balancing test' establishing the most appropriate level of data protection required in different circumstances when personal data is processed.

Ultimately, the authors think that this test can be found satisfied *de facto* because it is comprised in the text/substance of the GDPR itself, compliance with which means following a set of requirements and obligations that they describe as forming effectively a comprehensive "Data Protection Compliance Program" ('DPCP'). They therefore propose that organisations should be allowed to process personal data upon the condition that they have implemented (and can prove they have implemented) the DPCP – i.e. comply with the basic principle (Article 5) requirements of the GDPR. Additionally, and importantly, the authors argue that by organisations acting in this way the 'appropriate balance' (what they call data protection legitimacy) would be presumed, and it becomes unnecessary to satisfy "*more traditional legal bases of achieving legitimacy*". That is, a controller would not need to either obtain a data subject's consent to processing, or satisfy the legitimate interest lawful basis, in those circumstances where no other lawful bases are available to them as long as they have proof of compliance with the DPCP. This is because the authors believe that, as lawful bases, consent and legitimate interest "*do not necessarily offer an appropriate level of protection of personal data, nor do they support a flexible implementation of the user control paradigm"* (p.245).

Balboni et al's legal concept of a presumption of data processing legitimacy grounded on appropriate protection (that the data controller has fulfilled its duties and obligations of protecting data during processing in the circumstances) invites comparisons with the block exemption/safe harbour models introduced in the next chapter and outlined in more detail in Chapter 6 below. Model 2, in particular, is put forward by this thesis author as a way to use pre-determined regulatory presumptions to promote data protection outcomes that achieve an appropriate level of safeguarding against negative effects, and encourage beneficial effects, in different processing scenarios involving data relating to persons. Consequently, when compliance with a block exemption's criteria is achieved and can be demonstrated as such, data processing legitimacy should be presumed (with nothing more required). Like the DPCP model, therefore, it relies on a data controller's commitment to compliance (and demonstration of such compliance) with the relevant block exemptions duties and obligations, with automatic incentives available upon

compliance. Both models, therefore, build upon the strengths of the GDPR in setting forth a heightened level of accountability on controllers, by putting an emphasis on certain outcomes to be achieved in terms of good data protection governance, but also enhancing legal certainty.

Legal literature review summary

Emerging from the legal literature is a noticeable discontent with the legal concept of personal data/PII under data protection/privacy laws. Some of the legal literature (notably works by Gratton, and Hon et al) recognise that data protection needs to focus on addressing the underlying concerns that the concept was intended to avert. Yet scholars who include consideration of likely harm into models of personal data are few and, moreover, such models incorporating aspects similar to the Effects-based Approach and subsequent modelling proposed in this thesis are distinguishable in key ways.

This investigation of jurisdictional boundaries – providing more rigour in determinations of when data protection law rules apply and to what extent – may also be recognised as linked to determinations when data processing activities (i.e. usage of data and sharing arrangements for secondary reuse of data) are legitimised and indeed to be encouraged under data protection law where beneficial, which will be considered in Chapter 6 below. By refocusing data protection law jurisdiction on effects and by proposing a better framework for assessing the risk of privacy in that context – and a model (a block exemption/safe harbour regulatory/compliance 'tool') to effect that framework - this thesis will shows this paves the way to a future area of meaningful and necessary legal reform complementing but surpassing the GDPR. This is novel insofar scholarly and policy discussions have not examined explicitly this hypothesis and notably alleviates some of the weaknesses in other models.

4.3.2 Critical assessment of the Effects-based Approach in light of limitations that can be proposed to the sub-notion of 'Relevant Effect'

It is clear that there remain interpretive gaps to fill in developing scholarly interpretations reliant upon an effects-based perspective on the personal data concept and, specifically, around evolving the contours of the Effects-based Approach and its application to the 'grey areas' of identifiability per Chapter 3 (such as with respect to IP addresses and profiling activities). This section considers this point, while the next section assesses the efficacy of the model set against the twin aims.

As a reminder of the key thesis argument presented so far, under the Effects-based Approach model developed, data would be deemed personal data where a 'Relevant Effect' from a particular processing activity upon an individual (to whom it has already been established that the data would

be related) is likely. Said otherwise, the same test would need applying in respect of further (and all supplementary) proposed data type activities by the data holder on data that relates to the same individual (as well as in respect of data that relates to other individuals).

Already alluded to is one limitational factor that would inhere in the concept of 'Relevant Effect'. This is the proposition that the magnitude (or severity) of negative effect(s) predicted to result from the processing activity under consideration must not be a 'de minimis' one (i.e. it must be at least non-negligible - said otherwise, likely giving rise to effects to an 'appreciable' extent). To consider this aspect, and develop the approach further, it is useful to break down the theory's dissection into four question-headings for analysis. These represent inter-related aspects related to the notion of 'Relevant Effect' and related risk-assessment factors for evaluation as proposed under the Effects-based Approach:[372]

- **Relevant Effect – 'how'?**
- **Relevant Effect – 'from whose perspective'?**
- **Relevant Effect – 'with what magnitude'?**
- **Relevant Effect – 'with what likeliness of occurring'?**
- **Relevant Effect – on whom?**

4.3.2.1 Relevant Effect – how?

Negative effect(s) is a broad concept potentially covering a vast spectrum of things. At the very least in the present context, it denotes a requirement that an activity of processing of data relating to an individual be likely to result in some form of adverse impact being suffered by that subject.

For possible guidance on how – in what ways - effects might be deemed relevant (in aiming to retain some consistency with underlying existing doctrine), we return to WP248 and WP218, respectively. The former says DPIAs focus on risks "*to the rights of the data subjects*", with the notion of 'risk' described as *"a scenario describing an event and its consequences, estimated in terms of severity and likelihood"*.[373] The latter opinion – on a risk-based approach in data protection law - is more precise, saying that it is *"related to potential negative impact on the data subject's rights, freedoms and interests"* and *"primarily concerns the right to privacy but may also involve other fundamental*

[372] Indeed, three (the first, third, and fourth) of the headings listed below are of the type that the WP has directly recognised as elements for taking into account in assessing the impact of personal data processing on data subjects (in Article 29 Working Party, 2014, WP217, p.36), described as: "*potential sources of risk that may lead to impact on the individuals concerned, the severity of any impacts on the individuals concerned and the likelihood of such impacts materialising".* See also p.38, in which the WP reiterates focus on: "*the likelihood that the risk materializes on the one hand, and the severity of the consequences on the other hand - each contribute to the overall assessment of the potential impact".*

[373] WP248, p.15: "*the DPIA under the GDPR is a tool for managing risks to the rights of the data subjects*".

rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion".[374] In other words, reflecting upon the ways in which negative effects may flow consequent to a processing activity leads us naturally, first, to asking about the types of individual (data subject-related) interests potentially affected that may be deemed worthy of protection under data protection law.

The most likely issue of relevance (for determining the contours of 'Relevant Effect') are negative effects on privacy interests.[375] The protection of privacy interests relates closely to the protection of data protection interests, at least in the former's informational variety. Legally, they intersect as evidenced by the jurisprudence of the CJEU where discussion of both often intertwines.[376] However, the exact legal relation between the two concepts (and the protection of other fundamental interests in law) remains unclear.[377] Furthermore, under the GDPR, negative impacts on privacy interest appears distinguished from negative impact on data protection interests as a matter of fact, as evidenced by the choice of phrase '*data protection* impact assessment' (rather than its 'privacy' counterpart).[378]

Nevertheless, in helping determine what types of effect should be relevant under the Effects-based Approach, such statements – referring to impact on privacy, or other, interests - are still very

[374] WP218, p.4. Compare Recital 52 GDPR: *"[t]he likelihood and severity of the risk should be determined by reference to the nature, scope, context and purposes of the processing. Risk should be evaluated on the basis of an objective assessment, through which it is established whether data-processing operations involve a high risk…of prejudice to the rights and freedoms of data subjects".*

[375] Continuing the quote cited in the footnote directly above, the WP goes on to say (in WP218, p.4) that such risks, *"should be determined taking into consideration specific objective criteria such as the nature of personal data (e.g. sensitive or not), the category of data subject (e.g. minor or not), the number of data subjects affected, and the purpose of the processing. The severity and the likelihood of the impacts on rights and freedoms of the data subject constitute elements to take into consideration to evaluate the risks for individual's privacy".*

[376] See, e.g. Joined Cases C-465/00, C-138/01 and C-130/01, Rechnungshof v. Osterreichischer Rundfunk, [2003] ECR I-04989.

[377] Per Chapter 2, the protection of persons in relation to the processing of their personal data is a fundamental right laid down in the Charter. However, in some of its judgements, the CJEU seems to grant it an ancillary/instrumental status to other fundamental rights (especially, the right to respect for one's private life). In the *Lindqvist* case, for example, the CJEU emphasised that data protection's purpose is to establish a fair balance between *any* fundamental rights affected and the free flow of information. In other cases, the CJEU has been more inclined to underline the independence of the right to personal data protection. For example, in Case C-543/09, Deutsche Telekom AG v Bundesrepublik Deutschland [2011] ECR I-03441 (para.50), the CJEU declared that the DPD was "*designed to ensure, in the [MSs], observance of the right to the protection of personal data*"; thus, there was an omission of reference to other rights.

[378] See Gellert & Gutwirth (2013, p.529): *"[a]s a matter of fact, the European Commission's proposal for a general data protection regulation (GDPR) seems to favour a more autonomous version of the right to personal data protection, though it refers to a wording that is ambiguous. Such theoretical developments have been (at least implicitly) echoed in the discussion concerning data protection impact assessments (DPIA) and data protection by design (provided by Art. 35 and Art. 28 of the GDPR). Whereas in other legal cultures these tools are coined as "privacy impact assessments" and "privacy by design", the proposed [GDPR] uses the data protection vocabulary. Discussions have thus gone on to question whether a DPIA will mean assessing the compatibility of a proposed data processing operation not only with the provisions of the directive, or if it also entails to assess compatibility with the right to privacy (and eventually the whole spectrum of human rights)?"*

abstract.[379] Also very abstract are underlying concepts, such as the privacy concept, much discussed and debated, for which there is no universally accepted meaning or unequivocal definition.[380]

While defining privacy is the subject of a great deal of literature, there is no room in this thesis to describe at length theories of privacy definition. Instead, it is noteworthy that challenges in defining privacy as a concept include the fact that cultural views/perceptions about privacy differ worldwide,[381] plus it encompasses many different forms (physical and non-physical).[382] Notwithstanding, a common theme in scholarly discussions on the essence of the privacy concept is individual autonomy.[383] In particular, many scholars equate informational privacy with the ability to control the flow of information or knowledge about oneself.[384]

---

[379] They also may lead into dangerous territory of creating a tautologous logic in answering the overall thesis research question. In other words, the following proposition is only superficially true: the best model for achieving the objective of achieving a high standard of privacy protection via the legal concept of personal data (which data protection law seeks to regulate when it is processed) requires anchoring a new definition to the protection of privacy rights of individuals to whom such data relates and situations where they might otherwise be diminished. Yet, coming up with a new model of personal data under data protection law necessarily requires answering the theoretical questions, 'of what type, and how strong, should the link between data protection and privacy protection be in law?'

[380] Solove (2002, p.1088): "*[c]urrently, privacy is a sweeping concept, encompassing (among other things) freedom of thought, control over one's body, solitude in one's home, control over information about oneself, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations. Time and again philosophers, legal theorists, and jurists have lamented the great difficulty in reaching a satisfying conception of privacy.*'" Solove (2007, p.756) has also argued, "*privacy is not reducible to a singular essence; it is a plurality of different things that do not share one element in common but that nevertheless bear a resemblance to each other*." See also ibid, p.760: "*[t]he term 'privacy' is best used as a shorthand umbrella term for a related web of things. Beyond this kind of a use, the term "privacy" has little purpose. In fact, it can obfuscate more than clarify*". Solove (2008) also argues that no single definition can be workable, but rather that there are multiple forms of privacy, related to one another by family resemblances. For that reason, he says scholars, activists, and policymakers have struggled to define privacy, with many conceding that the task is virtually impossible. He then offers a comprehensive overview of the difficulties involved in discussions of privacy.

[381] Marx (2015, p.119): "*[p]rivacy is a multidimensional concept whose contours are often ill defined, contested, negotiated, and fluid, depending on the context and culture".* Compare Buchholz (1992, p.118): "*[v]alues are different between people and reflect individual desires and beliefs. Values are properties that human beings associate with or assign to certain forms of human behavior, institutions, or material goods and services*".

[382] For example, Banisar & Davies (2000) have suggested that privacy is capable of taking a number of forms: information privacy, bodily privacy, privacy of communications and territorial privacy.

[383] To give but two scholarly examples, Kupritz (1998) identifies three central themes of privacy: retreat from people, control over information and regulation of interaction. Bygrave (2010)**,** by comparison, suggests that there are four principal ways of defining privacy: non-interference; limited accessibility; information control, and definitions that incorporate various elements of the other three, but linking privacy exclusively to intimate or sensitive aspects of personal life.

[384] For example, Westin (1967) espoused an informational-control conception of privacy. In particular, he defined privacy as "*the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"*. Compare, for example, Fried who defined privacy as "*control over knowledge about oneself*" (Fried, 1968, p.483), as well as Inness who described privacy as "*the state of possessing control over a realm of intimate decisions, which includes decisions about intimate access, intimate information and intimate actions*" (Inness, 1996, p.140). Similar definitions may also be linked to theories of informational self-determination: individuals should be able to decide for themselves the nature and extent of information disclosed about them, since control over disclosure identifying information is necessary for the development of autonomous individuals. For example, one description of informational (data) privacy from Elliot et al (2016, p.134) is "an *individual's freedom from excessive intrusion in the quest for information and an individual's ability to choose the extent and circumstances under which his or her beliefs, behaviours, opinions and attitudes will be shared with or withheld from others".*

Legally, the definition of privacy also remains nebulous.[385] Neither the DPD, nor the DPA, attempt its definition. More widely, different concepts of privacy are given prominence in different legal systems. For example, various international and regional human rights instruments acknowledge a right to privacy (albeit a non-absolute one).[386] In EU MSs' laws, however, there is often no single definition of the term 'privacy', nor agreement regarding what a 'right to privacy' might encompass. Additionally, different domestic legal systems disagree over where the outer limits of the private domain and that of the non-private (public) domain rightfully should be. For example, there are striking difference in emphasis between UK and other EU MSs' courts regarding the degree of protection accorded to private interests in law.

It is outside the scope of this thesis (and research question(s)) to look in detail at this issue, or come to firm conclusions on what the concept of privacy, or the right to privacy, or indeed the right to protection of personal data (as set out in the Charter), encompasses. Furthermore, there is the danger of the circularity of trying to give a definitive answer to the question of what is informational privacy where it is fashioned, "*not a value in itself, but the decisive factor [that] consists in the relation between a person and specific information*".[387] Said otherwise, defining privacy protection as that which is achieved through the regulation of the conditions under which personal data may be processed might effectively jettison its independent importance altogether in focusing squarely on the personal data concept instead. In fact, many scholars acknowledge that not every processing of personal data necessarily affects privacy, even if it is nonetheless subject to data protection legislation.[388] Thus, the scope of data protection may be seen as being both broader and narrower than that of privacy protection in law. This conclusion flows from the fact that data protection serves individual interests other than privacy, such as ensuring data accuracy and data quality. The protection of these other interests also needs reflecting in the personal data concept.

Yet, defining what data protection interests are exactly remains as hard as delineating precisely what privacy interests are.[389] Attempting to determine their essential characteristics leaves us with

---

[385] The view that pinpointing a legal definition of privacy is problematic explains why, traditionally, courts have shied away from defining it in an all-encompassing fashion.

[386] Per Chapter 2, privacy as an international fundamental human right is recognised in, e.g. the Universal Declaration of Human Rights 1948 (Article 12); the International Covenant on Civil and Political Rights (Article 17); the United Nations Convention on the Rights of the Child (Article 16); the International Covenant on Civil and Political Rights; and the International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (Article 14). In Europe, the most notable human rights legislative instruments are the ECHR and the Charter. In such instruments, however, privacy is far from being an absolute right. Instead, a balancing exercise is necessary between the value underpinning recognition of the right versus the legitimate interests of others and overriding public interests.

[387] Weber (2003, p.150).

[388] Phillips et al (2014). To note, also, claims that information is private often arise on an ex-post facto as opposed to on an ex-ante basis.

[389] Discussions on the same are echoed in respect of DPIAs in the GDPR, in terms of questioning what a DPIA will require in terms of assessing the compatibility of a proposed data processing operation with an impact on what interests. Discussions have thus gone on to question whether a DPIA will mean assessing the compatibility of a proposed data

indeterminate conclusions, not least in light of doctrinal uncertainties regarding the relationship of the right to privacy and the right to personal data protection (such as under the Charter) and, indeed, to what extent they are instrumental to the protection of other rights.[390] While addressing these issues is beyond the scope of this research (if, indeed, it were possible to reach any definitive resolution on these issues), notwithstanding, this is not considered fatal to the Effects-based Approach in framework form.

To move forward constructively with this debate, we turn again to the problem of contextuality. Solove, for example, proposes decoupling the concept of privacy from a fundamental human rights approach in preference for a concept understood in the context of solving certain problems.[391] By comparison, Nissenbaum has suggested the 'contextual integrity' model of privacy.[392] Such views do not aspire to provide a one-size-fits-all concept of privacy (or data protection), but instead posit the need for theoretical frameworks that take into account different factors requiring assessment in different contexts.

This shift in emphasis also prompts focus upon the capacity for intuitively-recognising instances of negative effects that flow from processing as relevant to the circumstances at issue. Even if privacy and data protection interests cannot be defined definitively, it does not matter as long as we are able to recognise in practice (and particularise likely exemplar-types of) negative effects flowing from processing activities upon data relating to individuals and justify them as such. Said otherwise, it is easier to find agreement in identifying harm that flows from particular processing of data relating to people, than it is to define the values underlying such harm in conceptual terms.[393]

This acknowledgement, however, begs another question: 'should 'Relevant Effect' encompass any types of harms to individuals that might result from processing activities?' This is another issue subject to much scholarly debate.[394] Van der Hoeven, for example, proposes a classification of four types of harm possibly arising because of the compromise of privacy protections in an informational

---

processing operation not only with the provisions of the GDPR, or if it also extends~~entails~~ to assessing compatibility with the right to privacy (and potentially other fundamental rights).

[390] See Gellert & Gutwirth (2013, pp.529-530), who argue that, under an instrumental conception, data protection may be seen as safeguarding not only privacy, but all fundamental rights - thus *"respecting data protection would entail respecting all the fundamental rights at stake".*

[391] Solove (2008, pp.101ff).

[392] Nissenbaum (2004, p.119).

[393] Hurley (2015, p.73): *"[w]hile passers-by on the sidewalk might be hard-pressed to give a textbook definition of privacy, they could easily provide several examples of variations of their privacy, together with severe real-world consequences of job loss, public humiliation, and damage reputation".* In order to better understand what protecting personal privacy can mean, Van den Hoven (1999) suggests that we should ask, 'what would we lost if we did not have privacy?'

[394] See, for example, in the US context, Finch (2014, The Evolving Nature of Consumer Privacy Harm, [online]). One of the papers mentioned is Calo (2011) discussing what it means to suffer a privacy harm.

context.[395] Yet, identifying the outer boundaries of the concept of harm that may result when privacy protections are removed or breached is not a straightforward exercise, even ignoring the fact that other types of harm resulting from processing going beyond privacy harms cannot be excluded from consideration, such as harms associated with data discrimination. Particular points for debate include how far to take into account types of harm that are non-economic, including emotionally-felt types of harm, such as a loss of self-determination or loss of trust.[396] What about indirect – and secondary causal - forms of harm flowing from a processing activity?[397] Where should the Effects-based Approach draw the line?

While acknowledging these debates, it can be acknowledged that it is not necessarily the case that the 'wheel needs to be reinvented' in this respect. We can find guidance in the GDPR on harm types in a data protection context. Recital 75 describes "*the risks to the rights and freedoms of natural persons, of varying likelihood and severity*" which may result from personal data processing as potentially leading, specifically, to the following types of "*physical, material or non-material damage*":

> [W]here the processing may give rise to 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; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data; where personal data are processed which reveal racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, and the processing of genetic data, data concerning health or data concerning sex life or criminal convictions and offences or related security measures; where personal aspects are evaluated, in particular analysing

---

[395] Van Den Hoven (2008, p.311): "*[i]**nformation based harm** – of which the obvious types are identity theft but which also may include harms to the person, which are only possible following the acquisition of data or information about the person. **Information inequality** – where information about purchases and preferences are used for the purpose of marketing, price discrimination without awareness on the part of the individual or being able to influence this process. **Information injustice** – where information presented in one context is used in another. **Restriction of moral autonomy** – where people are restricted or limited in their options for self-representation due to the omnipresence and pervasiveness of personal information. This may also be termed a restriction on the choices that the right to privacy protects".*

[396] Robinson et al (2009, p.49) point out: *"[i]dentifying damage or the resulting harm when privacy protections are removed or breached is a complex task. There may be direct and indirect forms of damage and they may have consequences upon the individual in a variety of ways, ranging from monetary to social, mental and physical".*

[397] Albeit, much law already deals with this issue. For example, as pointed out by Elliot et al (2016, p.65): *"Article 8 of the European Convention of Human Rights stipulates that everyone has the right to respect for his or her private and family life, home and correspondence. Article 12 of the Universal Declaration of Human Rights (1948) goes even further: 'No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.' The concept of 'a right to a private and family life' encompasses the importance of personal dignity and autonomy and the interaction a person has with others, both in private and in public".* Although, to note, claims in law that information processing has caused damage typically arise on an ex-post facto based.

> or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles; where personal data of vulnerable natural persons, in particular of children, are processed; or where processing involves a large amount of personal data and affects a large number of data subjects.[398]

By comparison, in WP203, the WP refers to consequences for taking into account in assessing processing impact as liable to vary in scale and scope, at the same time confirming their breadth as potentially including: "*potential future decisions or actions by third parties, and situations where the processing may lead to the exclusion or discrimination of individuals*"; moreover, *"[i]In addition to adverse outcomes that can be specifically foreseen, emotional impacts also need to be taken into account, such as the irritation, fear and distress that may result from a data subject losing control over personal information, or realising that it has been compromised"*. [399]

4.3.2.2 Relevant Effect - as conceived from whose perspective?

An associated challenge with determining what types of effects could be a 'Relevant Effect' under the Effects-based Approach comes down to perspective. Laws often recognise (rightly) that one can suffer harm in subjective ways.[400] Nevertheless, despite such recognition, which throws up its own challenges related to how to quantify non-monetary harm, there is another question at issue - 'who should decide what potential effects that may flow from data processing are relevant?'

---

[398] See also, e.g. Recital 85 GDPR, which introduces the notion of prompt notification to supervisory authorities about personal data security breaches in certain circumstances, referring to types of damage that might result to natural persons. These include, *"loss of control over their personal data or limitation of their rights, discrimination, identity theft or fraud, financial loss, unauthorised reversal of pseudonymisation, damage to reputation, loss of confidentiality of personal data protected by professional secrecy or any other significant economic or social disadvantage to the natural person concerned".*

[399] WP203, p.25. See also comments in WP217 (p.37) where the WP refers to possible consequences relevant in assessing the impact of data processing, including the following: "*potential future decisions or actions by third parties, and situations where the processing may lead to the exclusion of, or discrimination against, individuals, defamation, or more broadly, situations where there is a risk of damaging the reputation, negotiating power, or autonomy of the data subject. In addition to adverse outcomes that can be specifically foreseen, broader emotional impacts also need to be taken into account, such as the irritation, fear and distress that may result from a data subject losing control over personal information, or realising that it has been or may be misused or compromised, – for example through exposure on the internet. The chilling effect on protected behaviour, such as freedom of research or free speech, that may result from continuous monitoring/tracking, must also be given due consideration".* The WP also gives an example in fn.85 of the same document related to an example of possible consequences of a financially harmful type: "for ex*ample, if a data breach releases financial information that was meant to be in a secure environment, and this eventually leads to identity theft or other forms of fraud, or the risk of personal injury, pain, suffering and loss of amenity that might ultimately result from, for example, unauthorised alteration of medical records, and a subsequent mistreatment of a patient, must always be duly taken into account…"*

[400] Perhaps in reflection of the abstract nature of values upon which laws can be based. Compare Buchholz (1992, p.118): *"[v]alues are different between people and reflect individual desires and beliefs. Values are properties that human beings associate with or assign to certain forms of human behavior, institutions, or material goods and services*". Although not all laws recognise compensation for emotional distress.

Under an objective approach, it would be possible for data holders to identify certain types of effects that are likely to flow from specific processing activities that generally may be considered harmful to individuals. However, a third party may not be able to anticipate readily how data processing activities may harm a specific individual practically in specific circumstances. Thus, taking a subjective perspective, analyses of relevant circumstances may introduce different harms deemed by an individual likely to be felt personally (a personal judgement influenced by their feelings, tastes, and opinions).

So, adopting a subjective perspective for the ex-ante determination of Relevant Effect under the Effects-based Approach may inevitably require consulting with the individual first to whom the data relates about their perception of potential harm from data processing. This would clearly be unworkable.[401] So, what support is there for adopting an objective perspective for predicting likely negative effects, and what does this mean exactly?

Assessing an objective perspective on likely harm could involve deploying a practical test of whether effects upon the relevant individual are predictable such that a reasonable person of ordinary sensibilities would construe them as harmful. By comparison, McCullagh discusses the 'reasonable person' test in relation to the sensitive personal data concept:

> The reasonable person test is an objective legal test. Thus an organization needs to be able to demonstrate that it considered the circumstances around handling personal information and made a decision on what is reasonable in the circumstances.[402]

Under the Effects-based Approach, consideration of circumstances could similarly be based on the facts that a data holder could reasonably be expected to know, acting as a reasonable person. However, this still begs the question of what a reasonable person would deem harmful in particular circumstances.[403]

---

[401] To note, in WP248 – by comparison - the WP advises that data controllers should consult data subjects or their representatives on the intended processing as part of the pre-DPIA risk-assessment process (see Article 35(9) GDPR). Yet, it offers no guidance for determining when this consulting may be deemed appropriate. The WP only states (p.13) that controllers should document "*its justification for not seeking the views of data subjects, if it decides that this is not appropriate*", and (if the data controller's final decision differs from the views of the data subjects) *"its reasons for going ahead or not*".

[402] McCullagh (2007, p.200).

[403] It also relates back, more widely, to the issue discussed in the last section in respect of what harms types are formally acknowledged by society and when may it be said that new harms are so acknowledged. Compare McCullagh (2007, p.197) talking about data sensitivity: "*[f]rom a subjective viewpoint, the sensitivity of a particular value is derived through individuals making personal judgements. In contrast from an objective perspective, the sensitivity of a particular value is derived outside the personal experiences of those individuals faced with choices. In this situation values are part and parcel of the behaviour or object in question. A complex interaction between these two perspectives leads to the creation of commonly held societal values that are believed to produce desirable outcomes for society as a whole (Buchholz, 1992; Daleiden, 1990). The conflict between the objective and subjective viewpoints is resolved through the essence of public policy formulation process, i.e. negotiation and compromise (Rule, 1974; Sieghart, 1984)".*

Another way to approach this issue could require data controllers to take into account data subjects' reasonable expectations when assessing whether adverse effects are likely to flow from particular data processing activities.[404] Said otherwise, the question should be addressed: 'in the specific processing context, might some forms of harm be likely to arise precisely because it would be at odds with the data subject's (or a reasonable person in the data subject's situation) reasonable expectations for how data relating to them would be processed?' The nature of the relationship between the controller and the data subject, and how it might affect expectations about the way in which a piece of information about the latter might be processed, is relevant here.[405] Assessing this requires consideration of what would be non-customary and unexpected practice in the given context, and in the given relationship.

Thus, it may not be possible to formulate an entirely objective category of negative effect types under the Effects-based Approach; it assumes that harm is a function of the particularities of information-usage in context, rather than the type of information itself (i.e. it depends upon the relevant circumstances under assessment).[406] Yet an objective perspective of harm may still be possible even if certain harms could only be anticipated in the context of former interactions between the one who plans to process the data and the one about whom the data relates. For this reason, it is conceivable that assessment by the former of a particular data processing activity's

---

Notwithstanding, per the list in Recital 72 GDPR, many harm types related to information processing are intuitively recognisable.

[404] Compare, e.g., WP217 (p.50) which mentions data subjects' reasonable expectations ("*with regard to the use and disclosure of the data in the relevant context*") as a useful factor for assessing the impact of personal data processing under the DPD (in the context of examining the notion of legitimate interests of the controllers under its Article 7). It illustrates its point (at pp.39-40 of this document) by referencing, "*the fact that personal data is publicly available may be considered as a factor in the assessment, especially if the publication was carried out with a reasonable expectation of further use of the data for certain purposes (e.g. for purposes of research or for purposes related to transparency and accountability*)"; and, *"it is 'important to consider whether the status of the data controller, the nature of the relationship or the service provided, or the applicable legal or contractual obligations (or other promises made at the time of collection) could give rise to reasonable expectations of stricter confidentiality and stricter limitations on further use. In general, the more specific and restrictive the context of collection, the more limitations there are likely to be on use".* WP248 (p.9) also refers to the exceeding of the expectations of the data subject in the context of describing a criterion that should be considered as part of determining whether a set of processing operations requires a DPIA due to inherent high risk (where datasets "*have been matched or combined, for example originating 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*").

[405] Compare, WP203 (p.25), talking about further processing purposes and compatibility of purpose but arguably also relevant by analogy: *"[i]n assessing the impact of the further processing, …Relevant impact in a larger sense may also involve the way in which data are further processed: such as whether the data are processed by a different controller in another context with unknown consequences, whether the data are publicly disclosed or otherwise made accessible to a large number of persons, or whether large amounts of personal data are processed or combined with other data (e.g. In case of profiling, for commercial, law enforcement or other purposes), particularly if such operations were not foreseeable at the time of collection".* At p.24 of this Opinion, the WP also comments that – in considering what a reasonable person in a data subject's situation would expect his/her personal data to be used for based on the context of its collection – relevant factors include any legal statements made by the data controller at that time as well as what would be customary and generally accepted practice in the given context.

[406] Compare Booth et al (2004, p.15): *"whether specific data types may have such an effect can be assessed in a context independent fashion"*; *"[t]his is problematic … It would appear to presuppose that specific types of information may affect dissimilar individuals…in similar ways".*

capacity for negatively affecting the latter (taking into account the particular circumstances, including any expectations that a reasonable personal might have regarding how the relevant data would be processed) might be deemed a legitimate perspective for determining Relevant Effect.

4.3.2.3 Relevant Effect - with what magnitude?

Regarding effects quantifiability (see also the next sub-section), developing the Effects-based Approach requires theorising about degrees of harm likely to be suffered resulting from a particular processing activity in the relevant case-circumstances.[407] Indeed, the extent of the effect likely for it to be a Relevant Effect is a very important issue under the Effects-based Approach as a limitational factor, otherwise the criterion would not add anything over and above satisfaction of the 'result' element of the 'relating to' criterion currently in law.[408]

To reframe the issue, in order to apply the Effects-based Approach, the standard of magnitude of negative effects likely from data processing activities needs to be determinable and justifiable. What should be the likely severity standard of any resulting harm deemed sufficiently concerning that it justifies triggering data protection rules and why?[409] Said otherwise, to what extent should a processing activity being carried out on a piece of information be deemed likely sufficiently harmful to an individual to whom it relates for it to become personal data?

As suggested above, a 'de minimis' rule is proposed: the magnitude of negative effect predicted to result from the processing activity under consideration must at least be *appreciable* to be a Relevant Effect.[410] Oxford dictionaries define 'appreciable' as "*large or important enough to be noticed*",[411] while Collins dictionary defines the term as "*sufficient to be easily seen, measured or noticed*".[412]

---

[407] Compare, for example, the WP in WP248 (p.13): "*the selection of certain technical or organizational measures **may affect the severity** or likelihood **of the risks posed by the processing***" (emphasis added). See, also Article 24(1) GDPR setting out the basic responsibility of the controller in terms of complying with the GDPR: "*taking into account the nature, scope, context and purposes of processing **as well as the risks of** varying likelihood **and severity for the rights and freedoms of natural persons**, the controller shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation*" (emphasis added).

[408] In other words, it could lead to gross over-expansion of the legal framework.

[409] This question bears in mind the statement by the WP in WP203 (p.14): "*processing of personal data has an impact on individuals' fundamental rights in terms of privacy and data protection. This impact on the rights of individuals must necessarily be accompanied by a limitation of the use that can be made of the data, and therefore by a limitation of purpose".* Also, in WP203, the WP said that factors to take into account in determining purpose compatibility include the impact of the further processing on the data subjects (p.25-26), and the safeguards adopted by the controller to ensure fair processing and to prevent any undue impact on the data subjects. In this context, it also refers to taking into account the availability of alternative methods to achieve the objectives pursued by the controller, with less negative impact for the data subject, as a relevant consideration.

[410] In comparison, the GDPR mentions processing categories of personal data that are liable to result in *high* risks to individuals' rights (Article 35 on DPIAs), while it is logical to assume that at least *some* risks to individuals' rights from data are compatible with its processing *not* being caught by data protection rules (rather than assuming a zero-risk standard).

[411] See, Oxford English Dictionaries [online].

[412] See, Collins Dictionary [online].

Thus, an appreciable effect is construable as an amount large enough to be important or clearly noticed (as opposed to a trivial, i.e. negligible, or de minimis effect).[413] Appreciable effect is chosen as the threshold, rather than zero effect, because it is assumed that the potential for negative effects upon an individual is present *whenever* information that can be linked to a specific individual under the 'relating to' criterion is processed. Therefore, using a zero effect threshold for personal data would be unrealistic and disproportionate, leading to gross over-expansion of the legal framework.[414]

The next question for addressing is 'how to assess this in practice'?[415] In other words, what factors would need analysing to determine whether – in the context of a particular planned processing activity – the potential impact magnitude (i.e. the severity of possible negative effects flowing from the processing) should be deemed of a 'likely appreciable' standard?[416] Chapter 6 considers this

---

[413] Compare comments by the ICO (2017, #1, p.19) that a 'significant' effect "*suggests some consequence that is more than trivial and potentially has an unfavourable outcome*" (an analysis of that document - entitled 'Feedback request – profiling and automated decision-making' - has been carried out by the thesis author in Knight (2017, ICO requests feedback on new data protection profiling provisions, [online])). However, the term 'significant' is rejected as the effect-based standard appropriate for the Effects-based Approach, as it imports a higher threshold (comparable to the GDPR-mandated DPIA requirement of "high risk") than the lower threshold of 'appreciability' more adeptly suggesting non-triviality.
It is noteworthy, by comparison on that point, the WP has stated (in (Article 29 Working Party, 2017, WP251, p.10) that *"[f]or data processing to significantly affect someone the effects of the processing must be more than trivial and must be sufficiently great or important to be worthy of attention*". This interpretation of a 'significant' effects standard in this sentence appears relatively low and more akin to a standard that might be associated with 'appreciable' effects. Notwithstanding, the WP goes on directly in WP251 to suggest a higher bar than non-trivial, by rephrasing it thus: "*[i]n other words, the decision must have the potential to significantly influence the circumstances, behaviour or choices of the individuals concerned. At its most extreme, the decision may lead to the exclusion or discrimination of individuals".* (An example is given by the WP at p.11 of automated decision-making that results in differential pricing as potentially having "*a significant effect if, for example, prohibitively high prices effectively bar someone from certain goods or services*").
The latter interpretation of 'significant' effects as implying an elevated threshold – going beyond a 'de minimis' effects-based standard - is also more in line with its reference to another effects-based standard that the WP has commented upon (in Article 29 Working Party, 2016, WP244, p.3) regarding personal data processing that "substantially affects" its data subject (as part of the test for determining whether "cross-border processing" exists under Article 4(23) GDPR). The WP refers here to *"the most relevant ordinary English meaning of 'substantial include; 'of ample or considerable amount or size; sizeable, fairly large', or 'having solid worth or value, of real significance; solid; weighty, important' (Oxford English Dictionary)".*

[414] In other words, if the standard was lower than 'appreciable' the evaluation process may be liable to indicate that data protection law should apply when in fact the possibility of unfair processing occurring is so low, and the risk of being caught by data protection rules so high (encompassing false positives), that it would deter processing unnecessarily. On the other hand, a higher-than-appreciable standard was rejected to avoid false negatives (i.e. a higher standard may be liable to indicate that data protection should not apply when in the fact the possibility of unfair processing occurring is sufficiently high enough to be concerning (or else leave individuals exposed)). Such economic-style cost-benefit analysis (impact assessment and compromise) often holds sway at the heart of any public policy formulation process.

[415] Compare, e.g. comments made by the WP (in WP217, p.38) referring to how data subject impact from personal data processing may be assessed, albeit that it only give examples of levels of likely severity: *"[an] element of the risk assessment is the severity of the consequences of a materialized risk. This severity can range from low levels (like the annoying need to enter again personal contact details lost by the data controller) to very high levels (like the loss of life when personal location patterns of protected individuals go into the hands of criminals or when power supply is remotely cut off through smart metering devices in critical weather or personal health conditions)… processing of personal data having an impact on a minority of data subjects - or even a single individual only - still requires a very careful analysis especially if such impact on each individual concerned is potentially significant".*

[416] While noting, per WP248, that there can be change of the risk posed by processing operations as circumstances change (thus, DPIAs must be periodically reviewed and the processes they assess, and the decisions taken

further. Suffice to acknowledge for now (as commented by the ICO) that it, "*may be useful to establish an external recognised standard to measure such effects, instead of simply relying upon the subjective view of the controller or the data subject*".[417]

4.3.2.4 Relevant Effect - with what likeliness of occurring?

The next challenge regarding quantification of effects is determining the degree of likelihood (probability) of appreciable effects occurring (because of a processing activity carried out upon data that relates to a particular person) under the Effects-based Approach. In other words, any prediction regarding the effect that a particular piece of information's processing might have on an individual – such that it should be deemed a Relevant Effect under the Effects-based Approach - also requires a determination as to the likelihood of its occurrence.[418]

Again, positing a very low (e.g. 'remote possibility' or 'may be the case that') - or very high - likelihood of harm occurring to an individual as the standard against which personal data is found under the Effects-based Approach could lead to gross over or under expansion of the legal framework. Somewhere in between is 'a realistic prospect', 'more likely than not' or 'on a balance of probabilities', standard (typically equated with '50/50').[419]

---

documented). It states: *"[r]isks can change as a result of change to one of the components of the processing operation (data, supporting assets, risk sources, potential impacts, threats, etc.) or because the context of the processing evolves (purpose, functionalities, etc.). Data processing systems can evolve quickly and new vulnerabilities can arise. Therefore it should be noted that the revision of a DPIA is not only useful for continuous improvement, but also critical to maintain the level of data protection in a changing environment over longer time. Finally, a DPIA may also become necessary because the organisational or societal context for the processing activity has changed, for example because the effects of certain automated decisions have become more significant".*

[417] From ICO (2017, p.7), in the context of a discussion about the concept of 'significant effects' that may result from profiling activities (Article 22 GDPR, see fn.325 above). Preceding this quote, the ICO refers to examples of significant effects in the GDPR as including automatic refusal of an online credit application, and e-recruiting practices without any human intervention. It then goes on to say – broadly - that its *"[i]nitial thoughts on other **significant effects** of profiling include processing that: causes damage, loss or distress to individuals; limits rights or denies an opportunity; affects individuals' health, well-being or peace of mind; affects individuals' financial or economic status or circumstances; leaves individuals open to discrimination or unfair treatment; involves the analysis of the special categories of personal or other intrusive data, particularly the personal data of children; causes, individuals to change their behaviour in a significant way; or has unlikely, unanticipated or unwanted consequences for individuals*" (emphasis added).

[418] Compare, for example, WP248 (p.13) where the WP comments: "*the selection of certain technical or organizational measures may affect the severity **or likelihood** of the risks posed by the processing*" (emphasis added). Compare Article 24(1) GDPR setting out the basic responsibility of the controller in terms of complying with the GDPR: "*taking into account the nature, scope, context and purposes of processing as well as the **risks of varying likelihood** and severity for the rights and freedoms of natural persons, the controller shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation*" (emphasis added). Likelihood of harm occurred might also be considered linked to the issue of where to set a limit upon the possible indirectness of harm considered. For example, should a data holder take into account a risk of indirect harm that might happen after successive time periods as a result of an initial decision not to deem data personal (with the result that data protection rules are not applied initially)? If so, this could result in a very high degree of likelihood that all data related to a particular person will end up being treated as personal data just in case indirect harm may ensue later on.

[419] A 'more likely than not' standard is in fact the standard of choice by the WP in guidance on assessing whether personal data processing "substantially affects" a data subject in WP244 (pp.3-4): "*[p]rocessing can be brought within the … definition if there is the likelihood of a substantial effect, not just an actual substantial effect. Note that 'likely to' does not mean that there is a remote possibility of a substantial effect. The substantial effect must be more likely than not".*

Alternatively, delineating the threshold as a standard of belief may provide a more useful metric of assessment – which could range from situations where a belief by the data holder may be no higher than 'more than fanciful' up to a standard of 'beyond reasonable doubt'. Slightly lower is a standard that requires that it is believed 'reasonably likely' that an event will occur. The latter standard seems preferable, bearing resemblance to the standard set out by the ICO in assessing identification risk – "*[i]n practical terms, the risk of (re) identification must be greater than remote and reasonably likely for information to be classed as personal data".* Moreover, it seems appropriate that more should be required than making an educated guess - there must be objective evidence to underpin the prediction of reasonable likelihood that an event will occur.

4.3.2.5 Relevant Effect – on whom?

So far, the focus has been on looking at the effects of processing on individuals. However, the Effects-based Approach need not only be concerned about effects on individuals but could also look at the possibility for effects on collective interests. These effects include potential for group harm (where a group may be seen to have a legitimate interest in the processing of data),[420] as well as potentially the impact on society at large (e.g. if intended processing could lead to a loss of public trust, e.g. following care.data described at fn.40 above).

There are particularly interesting issues relating to an approach that encompasses consideration on effects on (i.e. potential harm befalling to) groups, and the possible impact on the emerging concepts of 'group privacy' and 'group privacy rights' (i.e. where privacy as a group right is a right held by a group as a group rather than by its members severally). Authors such as Floridi have written on the broad nature of such challenges.[421]

The definition of personal data and its conceptual relationship to the potential for harm to group interests is surely one such challenge that needs tackling. Currently, the definition implicitly seems to seek an appropriate balance between the value of a dataset for analysis and the privacy interests of individual data subjects. However, the relevant concern may be illustrated by scenarios in which the processing of data which is not, and perhaps never has been, personal data under the existing definition may nevertheless result in harm being caused to groups of individuals based on one or more specific characteristics that members share.

---

[420] See, e.g. comments in WP251 (p.11): *"[p]rocessing 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".*

[421] See, e.g. Floridi et al (2016), which contains a series of articles highlighting issues associated with group privacy and new challenges of data technologies to such interests arising in response to specific challenges).

Modern data analytic techniques, in particular those involving data mining and profiling, cast a new light on the ways in which we conceptualise categories grouped algorithmically. Analytics can extract correlations from data in sometimes unexpected ways using non-defined parameters to sort and classify previously undiscovered patterns of similarities. Predictive inferences and other behavioural claims resulting from such analytic processes may be made about existing or newly identified (non-apparent) groups. The claims might in turn affect or harm these groups where decisions follow based on derived or inferred data. A particular interpretation may be inaccurate when it is based on data sets that are biased or non-representative in some way and yet this fact remains hidden.

Group-level harm may then follow from analysis of aggregated data where the 'aim' of the grouping was not to identify individuals, but groups are being treated in a certain way based on factors, identified by the analytics based on statistical assumptions drawn from relationships with other associations and groupings. For example, data processing actions regarding a particular genetic group may create discriminatory conclusions with implications for other genetic groups – such as genetic discrimination when it comes to setting life insurance rates expected to increase in the future with more widely-used tests for polygenic, multifactorial conditions.

The above combination of factors lends itself to a search for more flexible data protection instruments to protect groups (particularly those groups unaware of potential harm to their collective interests arising from particular processing activities). Indeed, there is danger in ignoring group-level harms by a false confidence in legal provisions (or anonymisation techniques) that are designed to give priority to the individual perspective alone.[422] Such dangers may only be understood by further research into how particular processing activities in context necessarily play a central role in creating different types of group harm that may be non-obvious.

To lay the foundations for this exercise, it would be necessary to reconceptualise the schema of harms that might arise from processing data relating to people - i.e. illuminating the problems that can arise as well as the mechanics of privacy harm creation at a collective level. This work may provide a starting point for future investigation, particularly around profiling and automated decision making issues, and would help assess the feasibility and practicalities of expanding privacy rights taking into account the risk of harms heightened by certain analytical processes and decisions along the data processing life cycle. In this context, adopting the Effects-based Approach could signify a move towards formulating an assessment mechanism to help protect collective interests where they are challenged by particular types of processing activities.

[422] Floridi (2014), pp.1-3. Floridi highlights concerns about underestimating the risks involved in opening aggregated data to public use in cases in which groups of people may still be easily targeted.

In later Chapters, there is further discussion of the ways in which an effects-centric exemption-based analysis may be developed into a regulatory mechanism that could – in a way that the current concept of personal data cannot – specifically include a consideration of groups. Chapter 6 describes how it could be used to help achieve balance between the minimisation of individual/collective harm to an acceptable level, while at the same time helping organisations become more aware of factors affecting the effects of their processing conceived broadly and help reap the maximisation of associated collective benefits related to the use of data about people.

4.3.3 Summary

To summarise, under the Effects-based Approach as developed, only if it is reasonably likely that appreciable negative effects (a Relevant Effect) will befall an individual resulting from a processing activity (to be carried out upon information relating to such individual) will such information be personal data legally. In other words, does a data holder have a reasonable belief (based on the information it knows from the perspective of what a reasonable person would construe harmful in the circumstances) that a particular type of processing it has planned for information that relates to an individual may cause appreciable harm to them? (An ex-post facto determination that appreciable harm occurred - consequent to a particular processing activity - would also be evidence that data relating to a living person should *retrospectively* be considered personal data in that context, e.g. by DPAs or courts (although evidence on causation would also have to be considered). Chapter 6 revisits this issue).

Notwithstanding, further research around delineating the contours of Relevant Effect in terms of how effects are construed, as well as objective determination factors around identification of effects, under the Effects-based Approach is clearly needed. As described, open-ended issues persist around what types of negative effects should be considered as relevant, as well as how to estimate objectively the likelihood of consequential harm, and the magnitude of the harm likely resulting. While regulator guidance – in particular, new guidance such as WP248 which aims to ensure consistency of interpretation of the common criteria on the methodology for carrying out impact assessments – has started addressing these questions, there is still a long way to go.[423] To

[423] See, for example, Knight (2017, New EU guidelines on data protection impact assessments, [online]): "…*I can't help feeling that we are still some way to understanding the 'science' behind the quantification of risk levels that sits behind the GDPR (e.g. when a risk to a data subject might jump from 'low' to 'high' risk, or vice versa), and there is more work to be done in this area to ensure harmonisation by delving the legal concepts that sit on top (e.g. 'risks', 'effects', 'harm') and their meta-concepts (a typology of data protection-centric risks and harms) in turn. This is notwithstanding the fact that the WP encourages the development of sector-specific DPIA frameworks to better address particular types of processing operations; and Article 35(3)-(4) GDPR require national data protection agencies to produce a blacklist and a whitelist of processing activities that should or should not be subject to an impact assessment (to be submitted to the European Data Protection Board to supersede the WP under the GDPR)".* As suggested in WP248 (p.10), *"[t]he criteria*

repeat the comment of the ICO above, it "*may be useful to establish an external recognised standard to measure [data processing] effects*".[424]

Another issue arising relates to scenarios where the magnitude of harm is predicted to be significant, indeed possibly extremely significant, yet the likelihood of harm occurring is less than 'reasonably likely'? For example, what if the planned processing activity gives rise to the exposure of very sensitive data relating to a particular person (such as a HIV test result) to a third party, but the assessment was that – in the circumstances - there was less than a reasonably likely chance of this occurring? Should the remote potential for significant harm befalling the relevant individual resulting, for example, from the unauthorised divulgence of this information, render such information personal data notwithstanding under the proposed Effects-Based approach?[425]

On reflection, the answer must be affirmative (at least at this moment without further research). That is, under the Effects-based Approach, it would be necessary to assume that any data falling within the Article 9 GDPR (currently, Article 8 DPD / s.2 DPA) sensitive category be personal data regardless of the processing context.[426] Otherwise, it could give rise to a situation where secondary harm may result (in a later period) to an individual precisely because sensitive personal data was not processed according to data protection law (in the original period).

However, this acknowledgement does not necessarily undermine the internal coherency of the Effects-based Approach; for example, its logic may be seen as part of the evaluation of where to set

---

set out above [for assessing when data processing is likely to result in a high risk] can help supervisory authorities to constitute such a list, potentially with more specific content added in time if appropriate".

[424] ICO (2017, p.7).

[425] Compare by analogy WP comments (in WP248, p.8) - regarding determining when processing operations of a particular type should be subject to DPIAs due to their inherent 'high' risk – that one criterion for consideration is the presence of sensitive personal data. This is described as including *"special categories of data as defined in Article 9 (for example information about individuals' political opinions), as well as personal data relating to criminal convictions or offences*" but also other features. The WP comments, *"[t]his criterion also includes data which may more generally be considered as increasing the possible risk to the rights and freedoms of individuals, such as electronic communication data, location data, financial data (that might be used for payment fraud). In this regard, whether the data has already been made publicly available by the data subject or by third parties may be relevant. The fact that personal data is publicly available may be considered as a factor in the assessment if the data was expected to be further used for certain purposes. This criterion may also include information processed by a natural person in the course of purely personal or household activity (such as cloud computing services for personal document management, email services, diaries, e-readers equipped with note-taking features, and various life-logging applications that may contain very personal information), whose disclosure or processing for any other purpose than household activities can be perceived as very intrusive".* To note, however, the WP also remarks in an accompanying footnote that, "*[n]onetheless, if sensitive data are not processed systematically and on a large scale,* ***their processing does not automatically present high risks for the rights and freedoms of data subjects. For example, a data controller organizing a corporate event, and would like to know therefore what kind of food his guests are allergic to, could process these sensitive data exceptionally and would not need to perform a DPIA****. Similarly, processing of special categories of data by a medical doctor in a one-person practice should not be considered "large scale" (recital 91)"* (emphasis added). Thus, the WP's advice is equivocal.

[426] This is because the processing of sensitive data is typically associated with causing significant (a severe level of) impact on the data subject.

a limit upon the possible indirectness of harm considered in assessing Relevant Effect.[427] Again, admittedly, this issue requires fuller examination to minimise uncertainties that are beyond the scope of this thesis' word count, although see discussion in the next chapter about such assumption (as an automatic legal presumption) possibly being capable of being overridden evidentially on the facts.

Yet, the Effects-based Approach does more than gain traction on, and suggest a theoretical model replacement for assessing, the intangible notion of identifiability. By data holders asking themselves questions about likely negative effects (impact/harm) from data processing right from the start, this will help them to formulate their own practical understandings of the need for data protection rules to apply. Thus, it has regulatory benefits (the next chapter discusses these further).

Separately, there are more practical issues still requiring resolution. How should data holders be guided in making factual findings that the 'Relevant Effect' standard is met in terms of an assessment framework/methodology? Can associated guidance contain illustrative analyses of possible harm types that could befall individuals as a result of data processing activities and when they might be considered 'appreciable', e.g. such that the type, extent, and likelihood of harm arising would not be ignored by a reasonable person? The next section explores these points further, as does Chapter 6 below.

In fact, however, we need not look too far already for part of a possible 'template' by taking inspiration from Recital 90 GDPR which contains the basic criteria needed to assess data protection impact posed by particular processing in overlap with well-defined components of risk-management.[428] Also relevant are WP217, WP248 and WP203 as each discusses components for assessing (and managing) the impact of envisaged processing operations on data subjects.[429] Parts

---

[427] For example, should a data holder take into account a risk of indirect harm that reasonably likely might happen after successive time periods as a result of an initial decision not to deem data personal (with the result that data protection rules are not applied initially)?

[428] Recital 90, GDPR: "*a data protection impact assessment should be carried out by the controller prior to the processing in order to assess the particular likelihood and severity of the high risk,* ***taking into account the nature, scope, context and purposes of the processing and the sources of the risk.*** *That impact assessment should include, in particular, the measures, safeguards and mechanisms envisaged for mitigating that risk, ensuring the protection of personal data and demonstrating compliance with this Regulation"* (emphasis added). Compare Recital 84 GDPR: "*[t]he outcome of the assessment should be taken into account when determining the appropriate measures to be taken in order to demonstrate that the processing of personal data complies with this Regulation*". While reiterating here the comments in fn.362 above that it is not being proposed that a full PIA/DPIA be carried out in assessing whether personal data exists. Rather, insights could be drawn from its methodology in determining the key factors for assessing Relevant Effects (such as whether the processing involves data that may be considered sensitive, the way that the data is to be processed, the purpose behind the data processing, the reasonable expectations of the data subject with regard to the use of the data in the relevant context, as well as measures envisaged to address identified risks of harm).

[429] As mentioned, in WP217 the WP sets out useful factors for assessing the impact of personal data processing under the DPD including (pp.50-51): the nature of the data, such as whether the processing involves data that may be considered sensitive or has been obtained from publicly available sources; the way data are being processed; the reasonable expectations of the data subject, especially with regard to the use and disclosure of the data in the relevant context; and, the status of the data subject and the data controller. By comparison, in WP248, the WP also specifies

of the UK Anonymisation Decision-Making Framework, for example, also highlight assessment of impact as one of the steps to be taken in assessing identification risk under existing law (see further discussion on anonymisation in Chapter 5).[430] Such guidance could be adapted and supplemented by more detailed practical guidelines on Relevant Effect. This is important not least because there is a need for read-across logic for individuals and organisations between the DPD/DPA/GDPR and the Effects-based Approach as a proposal for legislative change.

4.4 Critical analysis of the Effects-based Approach's compatibility with the twin aims of the DPD

In assessing the effectiveness of the Effects-based Approach (as a possible pre-requisite criterion to replace the identificatory requirement) under the legal definition of personal data – in terms of its potential for realising and reconciling the twin aims of data protection law - it is useful first to consider each aim in turn.

4.4.1 Protecting individuals

As the WP acknowledges, protecting the interests of individuals regarding the processing of data is of key importance when interpreting and determining how to apply the DPD's rules, not least because "*it may caution against any interpretation of the same rules that would leave individuals deprived of protection of their rights*".[431] This focus is also true of the GDPR, as borne out by its

certain features it considers could help assess the level of likely impact to result from a processing activity: a systematic description of the envisaged processing operations, the purposes of the processing; an assessment of the necessity and proportionality of the processing operations in relation to such purposes; an assessment of the risks to the rights and freedoms of data subjects; and, the measures envisaged to address such risks. (Annex 1 to these guidelines also contains a list of links to examples of existing DPIA frameworks (both generic and sector-specific) and to international standards containing DPIA methodologies for reference). Common to each method are three processes: establishing the risk-context; assessing the risks; and, mitigating such risks. This schematic is consistent with the accountability principle exhorted by the GDPR requiring controllers to demonstrate accountability through a genuine assessment of risk, not only to facilitate compliance with the GDPR, but also to demonstrate that appropriate measures have been taken to ensure compliance and document the features of their risk-assessments.

[430] Elliot et al (2016, p.117): *"[t]he five principles upon which the ADF is founded: …one aspect > Step 2: Assess the impact - Include in the plan how the potential impact might be assessed and recorded. The key questions here would be: … Are the data sensitive? … What is the nature of the harm likely to be experienced?"*

[431] WP136, p.4. Compare, ibid, p.25: "*[a]s a general consideration, it has been noted that* ***the European lawmaker intended to adopt a broad notion of personal data, but this notion is not unlimited.*** *It should always be kept in mind that the objective of the rules contained in the Directive is to protect the fundamental rights and freedoms of individuals, in particular their right to privacy, with regard to the processing of personal data.* ***These rules were therefore designed to apply to situations where the rights of individuals could be at risk and hence in need of protection. The scope of the data protection rules should not be overstretched, but unduly restricting the concept of personal data should also be avoided"*** (emphasis added). See also e.g., statements in a speech by the then European Data Protection Supervisor, Peter Hustinx (Protection of personal data on-line: the issue of IP addresses, 15 April 2009, [online], pp.3-4).

references to "*risks to the rights and freedoms of individuals*" (e.g. in Article 33,[432] and Recitals 76[433] and 77).[434] However, as mentioned earlier, what does "rights at risk" actually mean? There is a lack of particularisation in such statements around what such risks ('of what exactly?') might entail.

The Effects-based Approach steps in here at least with respect to the legal definition of personal data. It proposes to replace the assessment of risks of identifying individuals about whom data relates under existing/incoming law, with risks of adversely affecting them appreciably (resulting from a processing activity in respect of which such data *in that context* would be deemed personal data). Notwithstanding, as described in the last section, negative effects could potentially cover a vast spectrum of things, albeit it is acknowledged that the most salient type of effect under this model appears to be a negative privacy impact (via the likely materialisation of instances of privacy harm).[435]

To this end, it is the 'Relevant Effect' requirement – for assessing processing-context particularities in practice - that sidesteps the accusation that the argument is superficially tautological.[436] This requirement also avoids some of the identificatory-approach's problems by protecting explicitly individuals' interests (which the law specifically states it is there to protect when they are at risk) in terms of protecting individuals against harm that might be caused by them being affected negatively. It forces consideration of the overall objective of protecting individuals from harm where their rights/freedoms may be at risk by demanding a more explicit link between likely harm

---

[432] Article 33, GDPR: "*[n]otification of a personal data breach to the supervisory authority 1.In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent in accordance with Article 55**, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons**"* (emphasis added).

[433] Recital 76, GDPR: *"[t]he likelihood and severity of the risk to the rights and freedoms of the data subject should be determined by reference to the nature, scope, context and purposes of the **processing.** Risk should be evaluated on the basis of an objective assessment, by which it is established whether data processing operations involve a risk or a high risk"* (emphasis added).

[434] Recital 77, GDPR: *"[g]uidance on the implementation of appropriate measures and on the demonstration of compliance by the controller or the processor, **especially as regards the identification of the risk related to the processing, their assessment in terms of origin, nature, likelihood and severity, and the identification of best practices to mitigate the risk, could be provided in particular by means of approved codes of conduct, approved certifications, guidelines provided by the Board or indications provided by a data protection officer**. The Board may also issue guidelines on processing operations that are considered to be unlikely to result in a high risk to the rights and freedoms of natural persons and indicate what measures may be sufficient in such cases to address such risk"* (emphasis added).

[435] This admission does not preclude the possibility that the assessment of the negative effects likely to flow from the impingement of other rights and freedoms (relevant to the protection of data protection interests) may also be relevant for consideration under the model. Compare, Friedewald et al (2010, p.63"): "*Towards a new privacy framework - Data protection is both broader and more specific than the right to privacy. The relationship between these concepts is certainly something that needs to be addressed for a reconceptualization of privacy. **Data protection is broader because it not only aims at making the protection of privacy concrete,** but it also tends to protect other rights and interests such as freedom of expression, freedom of religion and conscience, the free flow of information and the principle of non-discrimination"* (emphasis added).

[436] In other words, maximising the achievement of privacy protection as a data protection law objective is easily done if the concept of personal data is defined in terms of privacy protection.

resulting from particular data processing and the personal data definition.[437] In so doing, it more adeptly frames the ex-ante risk-assessment that all those who plan to process data relating to persons must engage in already (the upfront jurisdictional analysis to determine whether planned processing will be subject to data protection law). It is also a logical step doctrinally. The Effects-based Approach could bring the two realms of legal protection (data protection and privacy protection) closer together.[438]

Notwithstanding, the premise that any information capable of identifying an individual is also thereby capable of having a Relevant Effect is not admitted. Determinations over whether personal data exists will not necessarily be the same in all circumstances applying models under the two approaches. Instead, the adoption of the Effects-based Approach could help close a gap of confusion that an identificatory-approach struggles to fill in relation to certain 'grey areas', by replacing it with a more logical decision-making framework. Moreover, the Effects-based Approach could better protect privacy interests than an identificatory-approach in such cases: a comprehensive approach should include considerations around data-usage explicitly as a key factor.

In illustration, we return to acknowledged data-type instances where data-identifiability is unclear. Notable is the case of dynamic IP addresses and related controversy over whether, and if so when, a particular person may be identifiable from such data (see previous chapters), especially where they are combined with additional information and intended for secondary-purpose (further) processing, all activities of which may lead to the creation of specific risks for related individuals. The Effects-based Approach seems more adept legally in clarifying whether data protection applies in such cases because it can take into account the processing activity under consideration on an iterative basis. Said otherwise, dynamic IP addresses would become personal data depending on

---

[437] As previously discussed, it may be argued that this is what the identificatory requirement, and identification risk analysis models, are really getting at. In essence, the fundamental concern relates to what might happen after identification disclosure from data, in terms of the potential for unfair processing affecting the individual to whom the data relates (such that it may be deemed sufficiently cogent to trigger the law's involvement), although this is rarely made explicit.

[438] Currently, it is acknowledged that privacy and data protection may overlap and apply to the same data processing situation, with data protection potentially conceived as both narrower and broader than privacy protection in law. It is narrower because it only deals with the processing of personal data, even if the latter does not infringe upon privacy, whereas the scope of data protection is wider because it potentially applies in situations not involving the processing of personal data, even if the latter does not infringe upon privacy. See Bellanova et al (2011, p.8): *"privacy and data protection, like other human rights, are legal instruments designed to safeguard the "political private sphere", but they have a very different mode of operation: privacy shields the individual, data protection controls and channels the instances that process personal data".* See also, Gellert & Gutwirth (2013, pp.522-523): *"we believe that privacy and data protection are products of distinct practices and 'regimes of enunciation', such as politics, law, ethics, economy, religion and so on, and that the challenge is not so much to find the foundational unity "behind" these, than it is to understand how, each being singular, they interact and articulate… the relation between data protection and other fundamental rights (and in particular the right to privacy) is far from crystal clear, as evidenced by Art. 1.1 of the directive"*. Notwithstanding, the Effects-based Approach is flexible as to the types of harms that it could take into account, as a matter of separate research.

the likely effects of the activity planned in terms of harm that may be suffered consequent to the processing activity under consideration. This seems a much more direct determination than under the skewered Means Test under Recitals 26 DPD/GDPR (liable to be mistaken in practice by data holders as requiring a one-off, static assessment of fact).

We can also consider data processing in the context of profiling more generally. Under the Effects-based Approach, the personal data concept would encompass any data relating to persons subject to profiling activities, regardless of identifiability, where appreciable harm is reasonably likely to arise for that person from its usage. Such an approach thwarts concerns around companies arguing that no-one is identifiable from data they use for profiling and, hence, data protection rules do not apply to its processing and further processing (with the added implication that this is a once-and-for-all-time organisational decision that they make at the outset only). The Effects-based Approach makes clear explicitly that there can be no assumption that any information might affect an individual in a relevant way because of its inherent properties a priori without considering the use of different types of information (with the exception of sensitive personal data – see above and the next chapter).

A new decisional framework is particularly important in a profiling-activities context because the provisions on automated processing decision-making (potentially including profiling activities, where likely to have certain significant consequences) contained in Article 22(1) GDPR can *only* apply where personal data are deemed to be processed in the first place. Under the Effects-based Approach, by focusing on the potential for a Relevant Effect to follow a processing activity – rather than identifiability - a high level of privacy protection is more likely, in the sense that the outcome determination would be more proportionate to (i.e. nuanced to take into account) the protective capacity required in the processing circumstances.

4.4.2 Facilitating data movement

As mentioned, EU data protection law aims to regulate the processing of personal data in order to support the free flow of such information through harmonisation of rules and, in turn, to promote legal certainty across the EU. Yet promises of legal certainty under the Effects-based Approach would depend on enabling assessors to achieve a sufficiently high degree of rigour and objectivity in determining whether a Relevant Effect exists, such that others (including DPAs) would nearly always agree with this assessment.

First, as to the theoretical aspect of how to interpret Relevant Effect, Recital 75 GDPR describes the types of harm that can befall an individual under data protection law.[439] However, issues around determining harm magnitude still present challenges. These include pragmatic ones around how to evaluate whether a level of harm is likely appreciable in a particular data processing context – which, while the evaluation may not require precise quantification, should at least indicate factors of sufficiency to raise concerns - even if determining whether such harm is reasonably likely to arise in specific contexts should be fairly intuitive. As Raab points out in discussing the PIA concept, which appears equally applicable to the concept of harm and its interpretation, under data protection law:

> [T]his concept is itself the subject of widespread debate concerning its objective or subjective dimensions, and concerning the way in which risks can be measured in the case of privacy, in which the questions of conceptualization and measurement are very difficult….A further important issue is whether we are talking about objective or subjective dimensions of the concept of risk; or rather, the relationship between the two... even if we are unable to estimate different degrees and kinds of privacy risk, to say simply that 'cookies pose a threat to privacy', or that 'people's privacy is at risk if databases are inaccurate' – however true – does not get us closer to a more nuanced understanding of what is more serious and what is less serious, or what is perceived to be the case and why, on which to base regulatory policy.[440]

Clearly needed is development of coherent and academically-sound theories of harm to underpin the data protection regulatory system (like those that have evolved using economics in the context of competition law, as discussed in Chapter 6). Nevertheless, as Raab remarks, this should not be considered an alien task from a theoretical perspective:

> Debates about risk analysis in a wide range of fields – for example, nuclear hazards, traffic accidents, environmental pollution and many others – have tried to understand the

[439] Recital 75, GDPR: "*[t]he risk to the rights and freedoms of natural persons, of varying likelihood and severity, may result from personal data processing which could lead to physical, material or non-material damage, in particular: where the processing may give rise to 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; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data; where personal data are processed which reveal racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, and the processing of genetic data, data concerning health or data concerning sex life or criminal convictions and offences or related security measures; where personal aspects are evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles; where personal data of vulnerable natural persons, in particular of children, are processed; or where processing involves a large amount of personal data and affects a large number of data subjects".*

[440] Raab (2005, p.14).

> probabilities of the occurrence of events, the magnitudes of their consequences and the effects of human activity and responses to the risks involved in these phenomena. [441]

Second, issues of legal certainty in a practical sense also arise in respect of whether a Relevant Effect is reasonably likely to arise in the set of circumstances under consideration. Of course, information about the data-environment context is critical for performing this calculation, yet this fact also has certain associated disadvantages as it:

- Does not allow you to draw up a list of data types that will ['always' or 'never'] be personal data.
- Any information might affect an individual in a relevant way given idiosyncratic vulnerabilities to the acquisition and use of information of different types.
- Relies upon judgements about the likelihood of a particular piece of information having an effect upon individuals which may prove incorrect in specific cases.[442]

In summary, it lays the Effects-based Approach open to criticism that it would require a casuistic form of regulation, which is more complex and lengthy than the alternative. Furthermore, predicting that a Relevant Effect is liable to flow post processing activity on a particular piece of information upon individuals may also prove incorrect ex-post.

However, contextual assessment is necessary as argued above.[443] It is also unavoidable when any dynamic and usage-dependent definitional models of personal data are proposed. Thus, a dynamic and risk-based identificatory model shares this legal certainty practical problem.[444] Like the Effects-based Approach, the latter is also liable to divergent interpretations of its theoretical components by courts and DPAs on the facts of particular cases.[445]

Notwithstanding, a test that relies upon predicting likely harm to individuals appears more intuitive than determining whether someone is identifiable from data taking into account additional information not in the hands of the one assessing such risk. Besides, considering the risk of harm in the context of a specific processing operation may be considered no more difficult than determining

---

[441] Ibid.
[442] Booth et al (2004, p.117).
[443] A non-contextual assessment would be problematic (ibid, p.15): "*because it relies upon an apparently untenable concept of privacy. It would appear to presuppose that specific types of information may affect dissimilar individuals' privacy in similar ways. This appears inconsistent with the concept of privacy familiar within either the sociological, psychological, or even, the legal perspective".* In other words, the Effects-based Approach acknowledges the significance of context to the assessment of whether a particular piece of data may harm an individual appreciably. See also ibid, p.16: "*context considerations are to some extent unavoidable when assessing an individual's privacy; it again has the potential to include all data; it relies upon judgments about the likelihood of a particular piece of information having an effect upon individuals which may prove incorrect in specific cases".*
[444] In fact, all ex-ante risk-assessment models are inherently fallible to some extent, being predictive and fact-specific.
[445] The difference is the extent to which each approach (and implementing models) can *limit* the possibility of significant differences in interpretation of the concept of personal data, effectively reducing legal uncertainty in practice. In other words, per fn.243 above, incongruity is caused mainly by the lack of a clear and coherent theoretical framework, which governs how context-dependent assessments should be carried out and aids understanding.

whether particular information is personal data by considering the risk of identification from that information:

> Whether information is 'personal data' already requires consideration of specific circumstances. Assessing risks of identification/harm posed by a particular processing operation would not seem more difficult than determining whether particular information is 'personal', yet may be more successful in making controllers consider the underlying objective: protecting privacy.[446]

We have some way, however, before it can be argued that the adoption of the Effects-based Approach would promote significantly more legal certainty because it is a much more practical and measurable test than whether someone is identified or Recital 26-identifiable from data. Yet, as mentioned, the Effects-based Approach has at least one clear advantage over the identificatory model assessed against the objective of maximising legal certainty: regarding the accretion problem (per Chapter 1). It is inherent to the Effects-based Approach that it requires retesting before each further processing activity involving the piece of information at issue. For example, a Relevant Effect may not be deemed to arise upon the creation of personal profiles, but it might be so deemed in respect of later use of such profiles. Precisely because it requires assessment in respect of each processing activity applied to data relating to a (living) person, the test is better than trying to establish the point at which, e.g., dynamic IP addresses become personal data because of the sufficiency and availability of other information from which identifiability can become a possibility.

The Effects-based Approach also better equates with acknowledgement that personal data status is not an attribute granted once-and-for-all-time, but is a characteristic determined by the context of the intended use that therefore has in principle to be reappraised as contexts change. Of course, this leaves open the prospect that assessing the status of a particular piece of information as personal in one processing activity context can co-exist with the possibility of classifying the same piece of data as non-personal data (or, indeed, personal data in contiguity with the original classification) in another (subsequent) processing activity context. Chapter 6 further addresses this point.

While iterative assessments under the Effects-based Approach might seem extremely burdensome, a single assessment may be deemed adequate to address a set of similar processing operations. Comparisons can be drawn here with Article 35(1) GDPR and its reference to a single DPIA being

[446] Hon et al (2011, p.226).

able to address multiple processing operations if they present similar high risks.[447] Moreover, organisations should be able to foresee what they are going to do with data even in big data contexts as pointed out by the ICO, and therefore the consequences of such use should also be broadly anticipatable.[448]

Notwithstanding, as mentioned in the last section, an effects-based approach methodology of factors for evaluation in context to lighten this burden on organisations practically should be developed, with worked examples of how processing activity types are likely to be associated with certain types of effects upon individual. Such examples could take inspiration from those used already in existing PIA/DPIA frameworks of assessment.[449] In particular, building on the most valuable insights found in such materials could help organisations achieve a more nuanced understanding of what is more serious and what is less serious regarding likely impacts from processing, and why (see also Chapter 6).

---

[447] Compare WP248 referring to similar processing operations as those that are similar in terms of nature, scope, context, and purpose (p.6): "*[a] DPIA may concern a single data processing operation. However, Article 35(1) states that "a single assessment may address a set of similar processing operations that present similar high risks".* Recital 92 GDPR adds that *"there are circumstances under which it may be reasonable and economical for the subject of a data protection impact assessment to be broader than a single project, for example where public authorities or bodies intend to establish a common application or processing platform or where several controllers plan to introduce a common application or processing environment across an industry sector or segment or for a widely used horizontal activity". This means that a single DPIA could be used to assess multiple processing operations that are similar in terms of the risks presented, provided adequate consideration is given to the specific nature, scope, context and purposes of the processing. This might mean where similar technology is used to collect the same sort of data for the same purposes. For example, a group of municipal authorities that are each setting up a similar CCTV system could carry out a single DPIA covering the processing by these separate controllers, or a railway operator (single controller) could cover video surveillance in all its train stations with one DPIA".*

[448] ICO (2014, #1, para 115, pp.36-37): ***"Unforeseen purposes …****The ability to analyse data for different purposes, such as using the location of mobile phones to plot movements of people or traffic is an important characteristic - and a benefit - of big data analytics... If an organisation has collected personal data for one purpose and then starts to use that personal data for a completely different purpose, it needs to update its privacy notice accordingly and ensure that people are aware of this. Furthermore, the idea that in a big data context it is not possible to tell people about the possible uses of their data needs to be challenged.* ***In general terms, big data analytics allows data to be used in innovative ways, but this does not mean that an organisation cannot foresee what use it is going to make of that data, and tell people about it****"* (emphasis added). Compare ICO (2017, #2, para 154, pp. 67-68). Moreover, data controllers have to know their personal data processing usages upfront under the GDPR, as well as the consequences of such processing in a profiling context. See, e.g. Recital 60 GDPR: "*[t]he principles of fair and transparent processing require that the data subject be informed of the existence of the processing operation and its purposes"*, and Recital 63: "*[a] data subject should have the right of access to personal data which have been collected concerning him or her, and to exercise that right easily and at reasonable intervals, in order to be aware of, and verify, the lawfulness of the processing. …* ***Every data subject should therefore have the right to know and obtain communication in particular with regard to the purposes for which the personal data are processed, where possible the period for which the personal data are processed, the recipients of the personal data, the logic involved in any automatic personal data processing and, at least when based on profiling, the consequences of such processing"*** (emphasis added).

[449] In other words, they currently grapple with similar type of theoretical problems outlined above, so developing the Effects-based Approach would not open a new debate as much as continue an existing one (including, admittedly, involving tricky issues) that currently tax, not just definitional but also substantive, data protection interpretations of primary legislation and associated concepts. Compare Elliot et al (2016, p.82): *"[t]hinking about the impact side of risk brings us to the third key concept, sensitivity, which tends to be connected with the potential harm of any confidentiality breach. However, as we will see, sensitivity is a larger concept than this and encompasses how the data were collected and what reasonable expectations a data subject might hold about what will happen to data about them"*.

For now, the links between the proposed evolution of the Effects-based Approach to enhance legal certainty, and its synchronisation with the major regulatory trends encouraged under the GDPR reform – such as impact assessments,[450] emphases on the accountability principle, and 'data protection by design and default', all requiring heightened responsibility by controllers[451] - are notable. Both put a focus upon data management via organisational self-assessments of likely consequences and the putting-in-place of mitigatory measures (to reduce the level of attendant risk so assessed) upfront. Also complementary is the fact that, under the GDPR, as under the DPD/DPA, there is expectation that those who process personal data should be aware of their intended processing-uses ex-ante (e.g., because they are expected to ensure *for all processing activities involving personal data* that they have legal bases justifying their future carrying out).

The Effects-based Approach would naturally fit into this schema and properly place the jurisdictional process of determining whether data protection law applies alongside the substantive process of determining how to manage the likely risks that might arise where there are plans to process personal data. Said otherwise, such measures all aim to get data holders to think about possible impact issues sooner, as (broadly) a single integrated compliance set of obligations of a similar type, which need regular review alongside other risk-management compliance reviews incumbent on 21st century organisations.

4.5 Case study applied to illustrate ideas discussed in this chapter

The discussion in this chapter about the Effects-based Approach and, in particular, about the notion of Relevant Effects is somewhat abstract. It would help to make this more concrete by illustrating the main ideas discussed in this chapter with an example using the facts of the (DeepMind) Decision case study introduced in Chapter 1 above. In particular, the question may be asked whether the alternative basis of appreciable effects or harm on data subjects be more satisfactory than the

[450] In Chapter 7, there is further discussion of a possible evolving theory of 'risks to right' emerging under data protection law and in particular under the GDPR, as discussed in Van Dijk et al (2016), and comparisons with the Effects-based Approach. This discussion pulls together some of the themes raised in this chapter about the value of assessing risks to the fundamental rights and interests of data subjects in carrying out obligations under substantive data protection law in the EU.

[451] Article 25(1) GDPR: "*[t]aking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects".*

current data protection regime's reliance on identification and identifiability as keystones of the definition of personal data.

In reminder, the Decision referred to the streaming of patient-related health data (current and historic) – encompassing over a million partial patient records - for the one-off testing by DeepMind of the clinical app, Streams, which was intended to be used for detecting and treating AKI. The ICO acknowledged that such partial records included personal data, such as names, addresses, health service ID numbers, photographs and videos. The patient was intended to be directly identifiable from such data: being information containing or intentionally linked to direct identifiers about patients (albeit encrypted in transit to DeepMind and at rest). Furthermore, the historical data could include very sensitive personal information held by the Trust regarding a patient's admission to a Trust hospital in the last five years. In other words, it was not in dispute that the DPA applied. Moreover, the Trust argued that it needed access to such real – and non-anonymised - patient data to demonstrate clinical safety before going live.

As to whether the relevant partial patient records would be considered personal data under the Effects-based Approach, this would require ex ante consideration – based on the evidence – whether a Relevant Effect (i.e. an appreciable negative effect reasonably likely to occur) would flow from the processing activity. In other words, the data controller (the Trust) would have to consider the potential for appreciable harm resulting to those individuals whose data was to be processed by DeepMind for the clinical safety testing phase of the project.

As mentioned, the controller would be aided with an authoritative (agency) methodology to guide it through this assessment process, setting out specific criteria to consider on the facts known to the controller at the time of the assessment. The methodology guidance suggested would aim to cover as many possible casuistics, including likely impacts on specific groups of individuals, although there may be very special cases that cannot be fully addressed under a general methodology. The controller would, of course, be able to take into account any mitigating controls it had already (or would) put in place to mitigate the likely to appreciable harm resulting from the processing activity.

In this instance, the Trust would be faced with consideration of a one-off clinical safety testing exercise where the objective harm resultant from the use of the data (by DeepMind) was low. As acknowledged by the ICO, no issues were raised about the safety of the data as there was evidence that explicit provision had been made for data encryption and security meeting industry standards (including on key management, with DeepMind not holding the decryption key).[452] Moreover,

[452] It is being assumed that this conclusion is correct according to high-level information security and information governance standards, and evidence of such was available to the Trust at time of assessment.

DeepMind was confined to acting under the strict direction of the Trust as processor as set out in the agreement between them. In particular, at the testing stage, DeepMind would not be engaged in any usage of the data other than to conduct clinical safety tests.

What about the likelihood of appreciable subjective harm that might (reasonably likely) flow from the processing of the data by DeepMind to the data subjects in this context, including potentially at a collective level? The issue is raised about whether there was adequate transparency to individuals whose data was used, including reassurance to them around adequate data governance practices in place. The processing of patient records by DeepMind for the testing of a new mobile app significantly differs from what data subjects might reasonably have expected to happen to their data after they attended the Trust hospitals within the last five years for treatment. On the other hand, the fact that the testing was a one-off occurrence might suggest that such subjective harm from usage of non-sensitive data would unlikely be appreciable.

Regarding the potential for collective harm, again touching on the issue of lack of adequate transparency (and in light of the failings of the care.data programme) - the large-scale data usage is notable. The disclosure of 1.6 million disclose patient records to test the app is of such a scale that it gives credence to the argument that public trust and confidence would be undermined (i.e. is likely to appreciably harm collective interests), particularly where there was a perception that information may be being shared for profit with a commercial third party organisation affiliated to Google.

In terms of potential positive effects at the collective level resulting from the processing, as mentioned in Chapter 1, the public benefits of testing/operationalising the Streams app are obvious. In later chapters, recognition of benefits such as these flowing from particular processing activities will be made explicit in the risk analysis (i.e. surpassing the consideration of implicit benefits from activities that facilitate data flows). It is worth acknowledging at this juncture that assessing benefits upfront can present similar problems to those discussed in this chapter around risk quantification (e.g. around the extent to which benefits have to be objectively demonstrable as likely to come into effect by the controller, and in what time frame). Therefore, a structured framework around assessing positive effects that may accrue (to individuals, groups, or to society more generally) from data processing activities is also introduced in that context.

Reverting to the question of whether the Effects-based Approach would be more satisfactory than the identificatory-based approach, under this case study it seems the outcome would be the same. Indeed, problems of legal certainty around determining a decision under the Effects-based Approach could put it at a disadvantage on the facts of this case, which in turn indicates the need for a strong 'tool' to require companies to account objectively for potential impact flowing from

their planned data processing. To this extent, a mandatory requirement to carry out a DPIA is a step in the right direction. However, it does not go far enough: this is because the GDPR requirement (Article 35) only applies after a controller has determined a) that personal data is going to be processed, and b) that such processing will give rise to a high risk. Something more is needed to facilitate an impact assessment upfront as part of the jurisdictional considerations about whether data protection law applies in the first place. As discussed in the next chapter where this case study is revisited with revised facts, this need is particularly relevant when data has been transformed by anonymisation techniques so the issue of jurisdiction (whether personal data are being processed) is more likely to be a point of contention. Next, we start to consider what this evolved effects-based model/tool might be.

4.6 Chapter conclusion

This chapter has addressed sub-research question 2 by way of an analysis of the sketched-out Effects-based Approach that can stand in contrast to the analysis carried out in Chapter 2 in respect of the identificatory-approach. In other words, an alternative approach is proposed from the status quo that would shift the emphasis to an assessment of the likely risk of harm to an individual and relevant others from processing.

It is suggested that the Effects-based Approach might become the only, or alternatively the predominant, criterion for determining whether data protection rules should apply to particular processing activities. This chapter tested the hypothesis that this approach is superior to the existing identificatory-approach in legislation (under which the application of data protection rules depends in most cases on a binary assessment of whether particular information is, or is not, to be classified as 'personal data' legally).

First, however, the chapter analysed critically other approaches to the concept of personal data that are possible – including those put forward by other scholars - and explained the reasons for their discounting in preference for developing the Effects-based Approach as proposed. It also demonstrated the compatibility of the theoretical emphasis of the latter with much existing doctrinal interpretations, so that essentially its adoption would retain (broadly) internal doctrinal consistency.

In addition, this chapter explored the benefits of the Effects-based Approach in, not only inciting a reflection on the shortcomings of the identificatory-approach, but also exposing what may be presumed to be the identificatory requirement's now-clouded instrumental purpose. In other words, the underlying concern signalled by this alarm may be interpreted as pointing to concern over the harmful effects that might otherwise result from the processing of data about particular

people in certain contexts. Arguably, the time has come to making a more explicit link between harm and the personal data concept, not least for its framing functionality in reminding data holders why the switching on of data protection requirements is needed for protecting individuals' interests when processing data relating to them.

Under the Effects-based Approach, a Relevant Effect standard has been proposed to provide the outer contours of the personal data concept in law in lieu of the identificatory requirement, alongside highlighted benefits regarding 'grey areas' where currently (and for the foreseeable future) uncertainty exists under an identificatory legal model of assessment. Therefore, the Effects-based Approach appears to offer more tailored protection of an individual's privacy than a model centred on a notion of identification, and may involve equal amounts of legal certainty as at present.

However, this acknowledgement belies the admission that any ex-ante (predictive and context-focused) assessment of negative events occurring in the future – including that which would be required if the Effects-based Approach were adopted into a legal definition of personal data - may be deemed dynamic and, consequently, fallible to inaccuracy viewed ex-post. Moreover, changing to a new legislative test would be costly and, therefore, must be justified as definitively superior to overcome resistance to change.

To improve legal certainty under the Effects-based Approach, the limitational parameters of Relevant Effect and how they should be interpreted theoretically and practically (especially how to deal with secondary, indirect notions of potential harm and unresolved related difficulties in determining whether sufficient 'appreciable' harm would be likely to arise) admittedly need further examination. For now, the suggested development of coherent theories of harm may sound idealistic, as importing an economics-based connotation not suitable to data protection/privacy type analysis. However, this is necessary to flesh out what is meant by references in the GDPR and elsewhere to differential risk-levels that individuals may face in relation to their rights/interests as potentially affected when data relating to them is processed. Furthermore, a more objective footing for the conception of harm in a data protection context would more easily allow discussions about degrees of risk, which is more practical in addressing concrete privacy issues pragmatically.

Yet – as demonstrated by the case study in the last section - it is difficult to avoid the initial conclusion that the Effects-based Approach so far sketched-out may well be less than effective in practice than anticipated. Without more, it shares many of the weaknesses of the effects-centric models described in the legal literature in this chapter (such as the model put forward by Gratton). This problem of outstanding weaknesses of the Effects-based Approach is addressed head-on in the next chapter and Chapter 6 below in exploring ways to mitigate this legal uncertainty through the

development of an effects-centric (block exemption/safe harbour regulatory) model. Arguably, this model would provide better ways to achieve the twin aims of data protection law – to more effectively reconcile them - compared to an identificatory-approach, but also the Effects-based Approach as presented in this chapter.

To further the debate in answering research sub-questions 1-2, the next chapter introduces the block exemption/safe harbour regulatory models and explores these questions from the new viewpoint that they bring: how to provide a practical guide for helping data holders when determining when information is *not* personal data after it has been subject to anonymisation techniques. This examination uncovers other 'grey areas' of legal confusion in the context of which the advantages and disadvantages of the two approaches – as developed in model forms - to personal data definition can be further compared. Specifically under consideration is the effectiveness of effects-based models in terms of providing regulatory incentives for minimising privacy-related risks when processing data following the application of anonymisation processes to it, while still permitting data value (utility) to be extracted from the further usage of such data post-transformation.

Chapter 5 - The concept of non-personal data and its implications for developing an effects-based exemption proposition

The chapter splits into two parts. **Part I** contains an interim thesis summary, concluding on sub-questions 1-2, to enable a preliminary answer to the overall research question. This answer is provided at this stage with a view to introducing the need for a significantly better effects-centric legal instrument (than the identificatory-approach, *or* amending the legal definition of personal data under the Effects-based Approach, can provide). This new instrument for proposal is in the form of an effects-based block exemption under the GDPR identificatory-approach framework. In particular, that answer signals the need to introduce sub-research 3, for discussion in more depth in Chapter 6.

As a reminder, the sub-research questions are

> **Under the existing identificatory-approach to personal data (exemplified by the statement, 'information can only be personal data if it is capable of identifying the individual to whom it relates'), how effective is this approach in terms of realising and reconciling the twin goals of facilitating the free flow of personal data, and safeguarding individual rights?**
>
> **Under an Effects-based Approach to personal data (exemplified by the statement, 'information can only be personal data if it is processed in a manner capable of affecting the individual to whom it relates appreciably'), how effective could this approach be in terms of realising and reconciling the twin goals of facilitating the free flow of personal data, and safeguarding individual rights?**
>
> **Can the practical disadvantages associated with an Effects-based Approach to personal data be ameliorated by the use of block exemption provisions, as exemplified under EU competition law (a distinct area of law and regulatory system that has been modernised using an effects-based policy approach)?**

Part II explores how existing EU/UK data protection law, and the GDPR, attempt to accommodate new technological complexities (per Chapter 1) under the identificatory-approach to the concept of *non*-personal data (i.e. data relating to living persons that is not personal data, for processing). This analysis considers the efficacy of a superior effects-centric legal interpretation of what it means

for personal data to be assumed no longer capable of identifying an individual post-data-modification, via the introduction of effects-centric legal instrument. This analysis focuses on what it means for a particular processing activity planned in respect of data relating to a person no longer to be deemed liable to be sufficiently impactful such that its processing falls within data protection law (whereas previously – pre-modification – it was so deemed liable).[453] This analysis is important for highlighting different perspectives on the benefits/disadvantages of each approach and, therefore, to help draw conclusions on which approach is adjudged better to achieve data protection's twin goals in balance.

In terms of chapter structure:

- **Part I** - **Section 5.1** concludes on sub-research questions 1-2, and provides a preliminary answer to the overall research question pointing to the need to address the final sub-research question the focal approach to which is scoped.
- **Part II** aims to present further evidence that there is room for an effects-based exemption legal instrument to sit alongside the GDPR pointing forward to Chapter 6. Section 5.2 recaps on current challenges related to anonymisation set against the backdrop of EU data protection law and evaluation of the identificatory-approach in that context; **Section 5.3** evaluates the prospect of an effects-based exemption in light of legislation and doctrinal interpretations aligned with an effects-centric analysis, as well as assessing the key strengths of a new exemption model in that context; **Section 5.4** concludes Part II.
- **Section 5.6** concludes the chapter.

Part I – Interim Summary pointing to the need for the Third Sub-Research Question

5.1 Why is a third sub-research question necessary?

This Part explains why the third sub-research question is necessary, indeed integral to the overall roundness of this thesis and its conclusions. It also signposts forward to later in this chapter and Chapter 6 in explaining the rationale for this question and the meaning of its key terms.

[453] In other words, in a near-identical context, the focal point of consideration relates to the lowering of the likelihood of data processing giving rise to a Relevant Effect through data-associated modification using anonymisation processes.

First, however, conclusions are drawn around sub-research questions 1-2 in the context of addressing the overall research question (piecing together the conclusions from Chapters 3 and 4, and anticipating some of the arguments to be made in Part II of this chapter below on the concept of non-personal data), thereby also highlighting thesis research gaps, some of which the addressing of sub-research question 3 helps fill.

5.1.1 Conclusion of analysis so far of the two approaches in the context of answering the overall research question at this stage of the analysis

It has been shown that there is still no uniform interpretation of the identificatory requirement of personal data in the EU under the DPD (with UK law used in illustration of just one MS's varied interpretational approaches to this issue under the DPD-implementing DPA). This fact gives rise to possibilities for inconsistent decisions about whether (and when) personal data exists in any particular circumstances. This fundamental incongruity has created, in practice, confusion that is likely to have lowered the protection of data subject rights, as well as possibly chilling the free flow of personal data because of significant legal uncertainty (as well as uneven enforcement) rather than harmonising data protection regulation pan-EU.

These challenges relate, to some extent, to establishing a clear rationale for the choice of words in Recital 26 DPD, confusion over how to assess the Means Test, and the extent to which residual risk is deemed acceptable outside the data protection regime. If we consider the concept of non-personal data, for example, the GDPR (and its Recital 26) provides some clarity to assist those who will need to assess re-identification risk (see Part II of this chapter for more discussion of this) in specific circumstances from 25 May 2018. For example, it refers to examples of "*all the objective factors*" that should be considered under the Means Test, such as the "*costs of and the amount of time required for identification*" (albeit that it is not clear, e.g., exactly whose effort and expense is being referred to).

Yet another open-ended issue of interpretation under the DPD persists, in relation to which the GDPR compounds uncertainty. For example, while Article 4(1) GDPR expressly refers to online identifiers, should this be interpreted as encompassing IP addresses? If yes, does this mean that the relevant part of the *Breyer* judgement will have to be ignored once the GDPR applies so that IP addresses should always be considered personal data? Alternatively, maybe this is not objectionable because of the GDPR's introduction if one were to presume that it could completely change the jurisprudence landscape (i.e. relevance of all preceding decisions under the DPD) in any event.

As seen from the case-study discussed in the last chapter, in comparing a dynamic (re-)identificatory risk-based model of personal data and the Effects-based Approach, admittedly the ultimate decisional outcome following application of each to particular circumstances might often be the same in practice. This will also be seen to be the case in many instances where it is planned to release anonymisation-modified data to a recipient).

This fact remains despite the fact the Effects-based Approach has advantages over an identificatory-approach, in particular:

- In previously-highlighted 'grey areas', where currently uncertainty and conflicting legal interpretations (often highly influential, but ultimately non-binding) of identifiability make it unclear exactly what data controllers must do in practice to be data protection law compliant.[454]

- The Effects-based approach could lead to a contraction, possibly substantial, of the personal data concept and, thus, help keep data protection law bounded because of the relevant criterion to be satisfied under the Effects-based Approach.

- The Effects-based Approach is coherent with many aspects of existing data protection rules and the GDPR. There is also policy support for an Effects-based Approach, most notably in ICO guidance, but also in some WP comments, focused on processing-usage (including dynamic context-dependent considerations around processing intended purpose and impact) as part of assessing identifiability. Indeed, the Effects-based Approach might even be considered intuitive and logically-appropriate in requiring upfront reflection on how data processing might affect those about whom the data relates, in keeping with the GDPR's emphasis on DPIAs and 'data protection by design and default'.

The weaknesses associated with the Effects-based Approach have been alluded to, specifically interpretational ones that would give rise to legal uncertainty. Notably, they include problems common to any dynamic model of the personal data concept reliant upon assessments of risk, reasonableness, and contextual-factor analysis to predict an outcome (compare Chapter 3).

Thus, the Effects-based Approach model as introduced in the last chapter should be rejected (alongside rejection of a proposal for a more extreme version of the Effects-based Approach under

[454] For example, as the GDPR does not reflect the WP's guidance (e.g. in WP216), it casts doubt on the applicability of the latter (particularly as neither the European Commission, Parliament, nor Council chose to follow the WP's guidance in its entirety).

which the concept of personal data is removed from the GDPR entirely,[455] or at least the removal of the legal concept of sensitive personal data/special category data[456]). This conclusion is supported by the fact that the probability of these proposals being implemented is admitted not to be realistic.

Nevertheless, there is room for future research around effects-centric legal models for discussion below.

5.1.2 Developing the thesis research path to answer the ultimate research question

Thus, while arguments have been set out for the superiority of an effects-based analysis applied to the definition of personal data over its identificatory counterpart, a third sub-research question is introduced to further research into alternative effects-centric legal models:

> **Can the practical disadvantages associated with an Effects-based Approach to personal data be ameliorated by the use of block exemption provisions, as exemplified under EU competition law (a distinct area of law and regulatory system that has been modernised using an effects-based policy approach)?**

We can restate such practical disadvantages as:

1) A paucity in legal certainty in understanding how to assess whether a Relevant Effect might follow a particular data processing activity; and,

2) Weak incentives for organisations to adopt an effects-based approach in assessing the personal data concept.

---

[455] As mentioned in the last chapter, such a position was advocated by Ohm (2010), who argues that privacy law should abandon the concept of PII. However, this would mean that data protection law would be left without a means for establishing a coherent boundary to necessary regulation. (In fact, Ohm goes on to propose using cost-benefit - risk versus utility derived from prospective information usage - analysis to determine how to regulate privacy, on the basis of quantified levels of re-identification risk and potential privacy harms assessed in specific contexts based on five factors (data handling-techniques, private versus public release, quantity, trust, and motive. Regulatory intervention, he argues (p.1768), should depend on the level of risk so assessed, such that regulators may choose to do nothing if the risk is very low).

[456] The main argument for removal of the legal concept of sensitive personal data is the fact that, per Chapter 4, data sensitivity is not a priori – it depends upon the contexts in which pieces of information relating to persons are processed. As the ICO has pointed out in its submission to the UK Ministry of Justice's call for evidence on current data protection law (IC, 2010, p.11), while many individuals would consider health data to be sensitive, *"is a record kept in a manager's file recording that an employee was absent from work because he or she had a cold particularly sensitive in any real sense*?"
Notwithstanding, in light of the generalised link between the processing of sensitive data and the risk of high impact befalling the individual to whom it relates, the thesis authors recommends retaining this formal legal category until a better approach can be proposed. Otherwise, the harm associated with false negative decisions could be substantial. See Elliot et al (2016, p.82): *"[t]he overarching point is that if you are dealing with sensitive data then the risk is higher both in terms of the likelihood of a deliberate attempt to access the data and the impact of an attempt if successful".* See also p.81: *"[s]ensitive data is thought to increase re-identification risk because (i) it is more likely to be targeted because it is interesting, and (ii) the impact (and potential harm) of a disclosure may be greater".* See also, e.g., Raab (2005, p.14): *"[t]he idea of 'sensitive data' implies that individuals would suffer greater damage through misuse of such data than they would through misuse of their name, address or other less 'sensitive' information".*

For discussion next is why these disadvantages are chosen specifically to help improve an effects-based model's overall effectiveness in achieving the twin aims. Specifically, addressing disadvantage type (2) may also help address disadvantage type (1), as explained in Chapter 6.

5.1.2.1 Improving legal certainty

The GDPR's emphasis on the accountability principle highlights the importance of improving legal certainty. Data controllers must demonstrate the carrying-out of necessary measures to ensure observation of substantive data principles/obligations upon their processing of personal data. This requirement also extends to considerations around possible obligations in processing data relating to data *possibly* personal, as well as producing paper-trails showing that necessary inter-/intra-organisational mechanisms for risk-assessment/mitigation have been implemented, to demonstrate compliance to external stakeholders (notably, DPAs) where needed. Small and medium-sized organisations without deep pockets to spend on legal advice, especially, need clear guidance on how to interpret legal terms in everyday situations.

A key consideration is the need for ongoing practical guidance in relation to an effects-centric assessment framework (especially regarding magnitude-of-harm certainty, and addressing theories of harm that arise with new processing type scenarios). For now, the suggested development of coherent theories of harm may sound idealistic, as importing an economics-based connotation not suitable to data protection/privacy type analysis. However, this is necessary to flesh out what is meant by references in the GDPR and elsewhere to differential risk-levels that individuals may face in relation to their rights/interests as potentially affected when data relating to them is processed. Decision-making methodologies by which organisations can identify and mitigate appreciable impact risks should be a focal-point, in a manner consistent with PIA/DPIA framework-of-assessment models and related authoritative guidelines.

5.1.2.2 Improving regulatory incentives

Improving incentives to undertake effects-based analyses in determining whether data protection law applies in the first place (i.e. at the jurisdictional point) – and also what obligations apply in particular - should also improve the level of rights protection.[457] These assessments will ultimately contribute to a more robust protection of personal data.

[457] Compare WP207 on open data and public sector information (PSI) reuse, which recommends that public bodies follow 'data protection by design and default' procedures (p.6). These include, as a matter of course, DPIAs carried out before such data relating to persons is made available for re-use, as well as the imposition of licence conditions specifically prohibiting the re-use of personal data for purposes that may affect the data subjects. Of course, different measures (and combinations of measures) may be appropriate depending on the processing activity and the other contextual factors.

Encouraging risk-mitigation to the point where the likely resulting harms from a processing activity is reduced to a level should result in lower privacy-risk processing generally. It also aligns with substantive compliance principles such as data minimisation and adequate data security.[458] For example, it would encompass consideration of harm potentially arising from external sources (such as risks of appreciable harm reasonably likely resulting to individuals because data relating to them is to be stored on unsafe servers).[459]

The importance of strengthening regulatory incentives also relates to the accountability principle, and the obligation to evidence the carrying out of on-going risk-assessments and data protection compliance, as data environments change. It also partners the acknowledgement that there can be no 'silver-bullet' solutions permitting data controllers to assume that data relating to persons can never be used in such ways that intended processing activities in the future might not harm them appreciably.

There is also a deeper point about regulatory incentives regarding the tendency of some organisations (e.g. online marketers) to argue that data they collect for processing are not personal data, whereas in fact it is only direct identifiers they have either *not* collected or have collected and later removed (but still leaving indirect identifiers of people in their possession).[460] Organisations acting cautiously might assume that personal data with direct identifiers removed remains personal data and, therefore, its processing must be in accordance with data protection rules. However, in reality, because the legal/regulatory onus (evidential burden of proof for discharging) falls on others - notably DPAs - to demonstrate non-compliance, organisations may act non-cautiously on the basis that this a 'grey area' of law. They may also consider it unlikely that they will be challenged because of the dwindling resources that agencies have to enforce data protection, and the high-cost of initiating private legal proceedings. Of course, organisations may not necessarily be flouting the law. Instead, an underlying associated problem is that organisations may assess (correctly) they are not subject to data protection rules at one point in time, but a changing data environment may mean that the same data later becomes personal data. Regulatory incentives may be considered off-skewed.

This problem will be addressed next with an effects-centric block exemption model proposed, which will be developed in Chapter 6.

---

[458] They both can require a range and combination of technical, legal, and organisational safeguards around disclosure control.

[459] In effect, consideration is required before storing any data relating to persons as to whether it should be deemed personal data – due to its sensitivity. This assessment should trigger steps to be taken to ensure that its security is adequate, otherwise, risk later (appreciable) harm occurring to the data subject through a data breach.

[460] See, in illustration, examples from the general review of Canadian industry privacy policies in Lo (2011), in particular p.26ff.

5.1.2.3 Proposal to introduce an effects-based exemption from identity-based data protection under the GDPR for assessment under research sub-question 3

As mentioned, given the fact that the identificatory-approach is adopted in the GDPR and is unlikely to be superseded in the near future, it may be that a proposal for an effects-based exemption from data protection rules is valuable as a legal instrument that would supplement the GDPR. In other words, this admits that, pragmatically, perhaps the theoretical exercise of forcing a stark binary choice between an identificatory-approach and an effects-one is inappropriate.

For example, it is acknowledged that existing assessment aids under the (re-)identificatory-approach (such as the motivated intruder test discussed in Part II of this chapter) retain value as useful proxies in assessing whether such data might cause harm to individuals post-disclosure. Not least this is because - while processing a piece of information relating to a person might not contribute to facilitating the identification of the individual – in practice, if someone (a data subject) could ***not*** at least be ***singled out*** from such information, ***then*** it may be difficult to sustain the claim that the information ***is*** personal under an effects-centric analysis[461] (setting aside collective harm issues for one moment).

This effects-based exemption proposal is akin to a 'half way house' model, which has been the preferred option of some scholars, notably Hon et al, with a focus on ex-ante risk of harm reduction in practice.[462] However, it must go further than them with additional benefits under offer as will be explained. In other words, might it be possible to make alternative effects-centric models better through remodelling (in particular, by considering an assessment framework that could be employed by data controllers when deciding whether personal data exists in particular circumstances)? Moreover, there is consideration of incentives and development of the argument that the a new model might provide additional benefits in encouraging data controllers to engage with jurisdictional issues in more than a perfunctory one-off manner (as is often the case now when it comes to considering data from which indirect identifiers have been removed).

---

[461] Indeed, this conclusion is unavoidable when an effects-based analysis is premised on the fact that harm must be likely to be befall at least the relevant data subject in isolation.

[462] As a reminder, they argue that – while the definition of personal data "*should be based on a realistic risk of identification*" – the "*applicability of data protection rules should be based on risk of harm and its likely severity*" (see fn.557-558 above). See, e.g. Hon et al (2011, p.46): *"[i]t is time to make the DPD fit for the next decade, and hopefully beyond…3. Anonymised data – consider clarifying the circumstances in which anonymisation may produce non-'personal data', with the likelihood of identification being the main determinant - for example treating as 'personal data' data which 'more likely than not' would lead to the identification of individuals. Data should not be treated as 'personal data' where there is insufficient realistic risk of identification. 4. Accountability – consider moving to a more nuanced, proportionate and flexible regime, which bases the general approach on end-to-end accountability rather than the binary 'controller/processor' distinction, basing the applicability of data protection obligations on the risk of harm and its likely severity, with appropriate exemptions. 5. Sensitive data – consider a similar risk of harm approach, with definitions as suggested by the UK ICO."*

Before moving to this analysis, three default ('analysis entry-point') presumptions are put forward flowing from the GDPR, which will highlight how a new effects-based exemption proposal could slot into that new legal framework:

- First, by retaining the current identificatory-approach under the GDPR, it is tantamount to accepting that in the future all data may be deemed personal (including information ostensibly not about individuals where a possible link to individuals could still be made at some point in time) and this could lead to gross-expansion of the legal framework. We would then need to learn how to embrace the principle that all processing of data should trigger the full remit of data protection law. In that context, introducing some legal mechanism for making data protection rules more scalable in practice is to be welcomed (otherwise risk having a disproportionate effect in diminishing the free flow of data in the EU). In other words, the block exemption idea would be valuable for keeping the scope of data protection within practical bounds, which would be a good thing.

- Second, a new model could effectively shift the legal/regulatory onus to fall upon organisations to demonstrate that any information relating to people under their control should not be treated as if it is personal data (and, therefore, its processing subject to the full gamut of GDPR rules).

- Third, an effects-based exemption may be considered a jurisdictional measure (even it did not involve the amendment of the personal data legal concept) because it could lead to a contraction, possibly substantial, of the personal data concept in practice.

To elucidate on this last point, the practical mechanics of how this would be achieved as a bolster to the GDPR regime - by advancing a step further in promoting its twin aims - will be outlined below and in the next chapter. For now, on a theoretical level, it is worth highlighting two points on how the success of data protection in the future must be grounded in a better understanding of the harmful effects of information processing:

- In re-emphasis, the emerging value placed upon data controllers assessing information-processing potential impact (including differentiating levels of potential impact, and implementing measures to lessen impact post-assessment) should be construed as equally important not just in relation to issues of data protection substance (e.g. when carrying out DPIAs made mandatory under the GDPR where personal data processed is deemed 'high risk'), but also to data protection jurisdiction. Consistent with understanding risk as falling on a continuum, this is because degrees of risk-mitigating safeguards for implementation (either through legal requirement or voluntarily to avoid legal requirement) also fall on a

continuum to match the gravity of risk so assessed. Both aim at the same outcome: that is, to reduce privacy risk, which can be equated with reducing risks of privacy harm resulting from processing activities under one's control.

- The argument that an exemptions-based proposal can sit happily alongside the GDPR and in harmony with it doctrinally (and the twin aims) can be supported upfront. Reference is made to the existence of exemption/derogation provisions already in place under the DPD and the GDPR, regarding restrictions and specific processing situations. Under Chapter IX GDPR (Articles 85-91), MSs have the option of introducing them for special purposes specific to their country (e.g. for the prevention and detection of crime, or for national security, or for scientific and historical research purposes or statistical purposes, etc.), as long as they respect the right to data protection and involve measures that are necessary and proportionate. Use of legal exemptions are, therefore, a valuable mechanisms for achieving better balance in appropriate data protection (between the protection of individuals with regard to the processing of their personal data and the free movement of such data, as well as the promotion of other welfare-enhancing pro-processing effects such as freedom of speech). They formally confirm the possibility of rule-flexibility in certain processing situations when less privacy protection might provide a better (overall) outcome than more. That might include application to those processing circumstances where a freer flow of information – promoting data reuse extracting hidden information value - would be likely to help to support knowledge-based economic activity (e.g. medical advances) to the wider public benefit.

Part II – Anonymisation Techniques and their Role in Legal Interpretation of the Concept of Non-Personal Data

Per Chapter 1, anonymisation techniques here mean processes applied to personal data. Data protection law acknowledges that personal data can be sufficiently transformed by such techniques that its further/secondary processing are no longer caught by its rules. According to Recitals 26 of the DPD and the GDPR, the principles of data protection shall not apply to personal data "*rendered anonymous*".[463] It is useful to consider why this is important in relation to two types of data scenarios.

One scenario is where an organisation applies techniques to data that – in its post-modification form – it intends to further process intra-organisation. The underlying desire here is often compliance-related. Upon further processing of personal data, organisations would find themselves subject to on-going compliance obligations, including finding a new legal basis for processing such data where the purpose is incompatible with the one for which is was originally collected.[464] This can be problematic practically as organisations often rely upon obtaining the consent of data subjects to the further processing of the personal data that relates to them but consent may not be forthcoming (and another legal basis not available, see fn.60 above and GDPR Article 6(1) for enumeration of the legal bases available in the alternative). Therefore, organisations in those circumstances have an incentive to ensure that the anonymisation techniques they apply to personal data have sufficient potency that the risk of not gaining such consent for further processing purposes (when it turns out that it was actually required) is limited as far as possible, or risk the consequences of getting it wrong (including the possibility of regulator fines).[465]

Another scenario involves data controllers intending to share the transformed data with a third party for further processing. Their motivation for applying anonymisation techniques is often also compliance-driven. If data protection rules apply to such further processing of data by third parties, the initial data controller would (rightly) be concerned about being held jointly liable consequent

[463] This assumption is often shared in other jurisdictions. For example, in the US, when organisations use data deemed to be adequately anonymised, their use of that data is typically not restricted under privacy laws or regulations. Notwithstanding, the terminology used can differ (see discussion next in Section 5.1).

[464] See Chapter 2 and its description of the purpose limitation principle found in Article 6(1)(b) DPD (implemented in the UK by paragraph 2 of Part I of Schedule 1 to the DPA) and in Article 5(1)(b) GDPR.

[465] Considerations of legal certainty are relevant here in that such organisations will hope that the relevant data protection authority with competence over their processing activities will share their views on the anonymisation standard in law that they aspire to (in the relevant jurisdiction). The legal onus is on DPAs (and courts) to conclude that, on the facts, the relevant data was still personally identifiable of the particular person to whom it relates before they can make a finding against the organisation regarding breach of data protection rules.

to third parties' actions if the latter breached data protection law.[466] Thus, anonymisation techniques are applied – as may be other measures with respect to the third parties' handling of the modified data - to ensure that the data shared is non-personal and alleviate that concern.

5.2 Recap of anonymisation-related legal concepts and their connection with the identificatory-approach

Traditionally, associations are made between the anonymisation concept and the identificatory-approach to the legal concept of personal data. These associations help explain why confusions are rife as to the exact consequences of certain anonymisation-related terminology in legal terms.

'Anonymisation' suggests cutting or obscuring the link between a person and a piece of information that identifies them in some way. In everyday speech, however, 'anonymised' data (personal data that has been subject to anonymisation techniques) is often equated with so-called 'anonymous' data from which it is no longer possible to identify an individual from that data now or in the future.[467] This is potentially misleading from a technical and legal perspective. From a technical viewpoint, per Chapter 1, scholars have argued that absolute (guaranteed) anonymity is impossible given today's powerful analytics, assuming that data value (specifically, attribute disclosure per Chapter 1) may still be extracted.[468] In this sense, anonymisation should not be considered a static end state, rather a fluid/dynamic concept to reflect the fact that the surrounding data environment is liable to constant flux.

High-profile cases are cited typically as justification of increased levels of concern about the effectiveness of anonymisation techniques.[469] Legally, however, an acceptable level of re-

[466] For example, data subjects may have been assured by the data controller that only 'anonymous' data would be published about them subsequently, which would be untrue if personal data collected from them was later published.
[467] To note, a distinction can be drawn between this concept of 'anonymous' data, and the use of the data-label 'anonymised' (sometimes used synonymously with the term 'de-identified' in certain jurisdictions, such as Australia). 'Anonymised' data, as the name indicates, suggests simply that it is personal information that has had anonymisation processes applied to it. In more sophisticated terms, and without making any guarantees as to zero re-identification risk, Bridges (2015, p.31) describes de-identification synonymously with anonymisation as *"one of the most common tools for protecting the privacy of data subjects…*t*he process of manipulating or transforming a data set such as to make it very difficult to discover a person's identity or attributes*". (In fact, in the UK, the term 'de-identification' can connote a process of removing only direct identifiers from personal data, in contrast to 'anonymisation' which alludes to a process whereby direct and indirect identifiers are removed. However, for the purposes of this thesis, that former connotation is covered by use of the term 'pseudonymisation' as discussed next).
[468] Compare, Knight et al (2014, p.7).
[469] Highly-publicised re-identification attacks that have occurred over the last decade are described by Ohm (2010), De Montjoye et al (2015), and Mantelero (2015). For example, in 2008, Netflix released 100 million film rental records after removing personal identifiers, as part of an attempt to improve its recommendation system. Researchers were able to re-identify particular users by comparing rankings and time stamps with public rankings and time stamps in the Internet Movie Database. More recently, scholars continue to proffer arguments and supporting evidence that it is still possible to identify individuals in datasets held out as being anonymous - see, e.g. Castro & Cavoukian (2014). Notwithstanding, such arguments are becoming increasingly complex to those without a technical/statistician background, as are the formal ripostes to such studies - see, e.g. Felten & Narayanan (2014).

identification risk from modified data - to avoid the application of data protection law in respect of the processing of such data - does not necessarily have to equate to zero re-identification risk.

5.2.1 Anonymisation techniques and re-identification risk

Traditionally, anonymisation techniques are considered to fall into two broad types. First, processes can be applied to remove direct identifying characteristics (e.g. name) from data; however, such redaction processes could still leave indirect identifiers or quasi-identifiers behind in the transformed data (per Chapter 1). Second, processes can be applied to personal datasets that aggregate individual data elements (e.g. by age group), such that summary level statistics may be produced that do not include individual-level data.

The focus of this analysis is on the first type, where data points remain related to individuals from which they can be singled-out, as it is in this scenario primarily that absolute anonymisation has become increasingly difficult to achieve and impossible to guarantee.[470] Yet, where individual-level data remains, this may be a desirable outcome for the data holder precisely because they, or a third party, may wish to cross-link information about the same individual represented by two or more data points (even if that individual's 'real-world' identity remains – or is presumed to remain – unknown to the data holder and/or the third party). Nevertheless, this residual ability raises the risk of re-identification. Specifically, additional information (from known or unknown sources) could become available facilitating re-identification by someone.[471]

5.2.2 Pseudonymisation and re-identification risk

Per Chapter 1, re-identifying someone from personal data (so-called 'de-anonymisation') has been described as, "*the process of determining the identity of an individual to whom a pseudonymised dataset relates*".[472] This quote implies that all data relating to particular persons that have been subject to anonymisation techniques from which, in turn, it is possible to re-identify such individuals, may be considered 'pseudonymised'. This implication derives from the fact that pseudonymised data is commonly agreed to be data that retain individual-level data, potentially enabling cross-linkability of additional information with such data. Whereas 'pseudonymisation' is

---

[470] Again, this is because of the increasing recognition that technological advances are making it increasingly simple to isolate individuals from datasets and connect them to knowledge through automated matching and pattern recognition, regardless of intent to specifically re-identify them (in a 'real-world' sense of the word). For more, see Ohm (2010). To note, this is not to deny that aggregated information (such as statistics) may still retain some risk of re-identification, where individuals can be identified because of the small number of individuals in particular fields of information and via data-outliers.

[471] Indeed, this source of information could be the data holder themselves who carried out the anonymisation techniques in order to share such data with third parties.

[472] Maude (2012, p.7).

commonly considered a data-transformative process, which involves deleting direct identifiers from personal data and replacing them with 'pseudonyms' (indirect/synthetic identifiers, such as numbers) potentially enabling such data cross-linkage.[473]

Per Chapter 3, the possibility to single-out individuals from pseudonyms in datasets suggests that data protection law *can* (and will continue to, under the GDPR) apply to the processing of pseudonymised data.[474] Said otherwise, the pseudonymisation process does *not guarantee* that data so subjected *cannot* be associated with an identifiable individual.[475] Notwithstanding, as mentioned, benefits can be derived from pseudonymisation because it allows a data holder or a third party to cross-link information about the same individual represented by two or more data points within a data set, or across two or more datasets. Moreover, in some cases data controllers may explicitly want to leave themselves open possibilities to retrace the 'real-world' identities of those individuals that are represented by pseudonyms (by retaining the raw/source data). This is often the case in health-related and pharmaceutical research, where personal data are often stored and processed in the form of 'key-coded' data.

In that context, re-identification risks, over and above the risk of an individual being singled-out from data discussed earlier, could arise in respect of three different sources:

- First, *data recipients* could match data accidentally within or across datasets and infer from that combined knowledge individuals' 'real-world' identities. Nevertheless, it is customary for data controllers to keep obvious information (e.g. code-keys) connecting the pseudonym to the data subject's 'real-world identity' separate via technical and organisational measures to inhibit their re-identification by data recipients.

- Second, some level of re-identification risk may still yet be deemed to arise by virtue of the fact that – assuming the original data are not destroyed – information retained by the *data*

[473] Although, to note, while this interpretation of the meaning of pseudonymisation is broadly mirrored by all others, the finer details of how pseudonymisation may be defined can vary. In particular, the formal definition of pseudonymisation in the GDPR discussed below provides a refined variety of this meaning.
[474] In other words, if a pseudonym is unique within a dataset or sets, the person represented by that pseudonym may be singled out from others represented within the same sets. In other words, this may permit matches to be made between the same person within the same dataset and/or across different sources of information. For example, data may be linked based on a common pseudonym in two databases, such as affiliation with a particular political party and an interest in nudist camping to person 'x' – and inferences drawn and decisions made based on such information – without 'x's 'real-world' identity being known or knowable via the raw data on which the pseudonyms are based (e.g. if irreversible encryption is applied to it).
[475] Compare, e.g., Hon et al (201, p.215) who state: "*[i]dentification numbers or similar unique identifiers may in particular enable linking of disparate information, associated with the same indirect identifier, to the same physical individual, to identify them*". Notwithstanding, this still leaves open questions around how the terms 'identifiable'/'identification' should be defined legally (per Chapter 3) and, in particular, the extent to which such terms are linked to a revelation (actual/potential) of 'real-world' identify as we commonly understand that turn of phrase. See the next section.

controller entity could (de facto) still enable the retracing of the data to the data subject's 'real-world' identity. In the context of health research, as mentioned, this possibility arises precisely by virtue of the common practice for controllers to retain the raw data and its code-keys listing those data subject names that correspond to the particular coded-pseudonyms used.

- Third, the code-key could be compromised and obtained by unknown *third parties* (even if it is perceived to be stored securely by the data controller).

5.2.3 Anonymisation under the EU data protection law framework

Unsurprisingly, there is much room for uncertainty over when data should be deemed non-personal data in law ("*anonymous information"* so described in in Recitals 26 DPD/GDPR[476]).[477] As non-personal data is the converse-concept of personal data, it also connects strongly with an identification risk-based approach (in this context, an approach that defines personal data assessed against the risk of re-identification from a piece of data relating to a particular individual).

Yet - from a factual viewpoint - the degree and source of re-identification risk from data transformed by anonymisation/pseudonymisation techniques can arise and overlap in many different ways. Perhaps most importantly, as mentioned, there is generally a trade-off between the amount of de-identification performed on personal data and the subsequent utility that can be extracted from the processing of such transformed data (linked to its transformed data-quality state). Thus, a risk-assessment must be carried out by data controllers as to whether an individual may become personally re-identifiable from modified data pursuant to the Means Test (Recital 26 DPD, or Recital 26 GDPR from 25 May 2018)).

It is generally accepted now that deleting direct identifiers from data is inadequate to prevent the presumption of personal identifiability from such data under EU and UK (as well as other MSs') data protection laws. As mentioned, endorsing the Means Test in relation to the concept of personal data suggests that EU law-makers did not anticipate or stipulate that the risk of re-identification from modified data should be zero. This approach is also reflected in at least some national laws implementing the DPD, albeit MSs vary in their interpretation of what constitutes an acceptable

[476] Recital 26 GDPR - similar to Recital 26 DPD – reads as follows: *"[t]he 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 to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."*
[477] To give just one upfront example, Beyleveld & Townend (2008, p.81) interpret the relevant standard for falling outside data protection rules as: "*where the data no longer has a history that can link it to an identifiable data controller who obtained the personal source data from the data subject or where it is known that the source data was given for unlimited purposes".*

level of residual re-identification risk assessed against this Means Test standard (hereafter, for brevity, termed a standard of '**Functional Legal Anonymisation'**).[478]

Thus, corollary questions arise regarding how much re-identification risk mitigation must be carried out - and in what ways - such that this standard is met. Said otherwise, the concept of non-personal data implies the denoting of a line between acceptable, and unacceptable (that is, unacceptable without legal protection applying to prevent breach of the data protection principles) levels of re-identification risk in a particular data processing context.

We examine these questions next, along with the advantages/disadvantages of adopting an identificatory-approach model underpinning the concept of non-personal data under data protection law, taking into account the existing legal and policy framework alongside recent developments in this area. Also considered are justifications given for deciding when sufficient efforts are likely expended by data controllers such that the further application of anonymisation techniques to data are considered unnecessary to avoid the application of data protection rules.

5.3 The identificatory-approach and the concept of non-personal data

5.3.1 Interpretations of the anonymisation concept and its relationship to re-identification risk under the DPD/DPA

Over the last decade in particular, EU data protection regulators have issued non-binding, yet persuasive in their legal authority, anonymisation guidance to assist data controllers in determining how the law applies in this area.[479] Two key ones especially provide sets of general principles and practical advice on methods for anonymising data that can be applied, as well as associated risks in publishing such data. Both also interpret how the DPD Means Test should apply when assessing data that has been subject to anonymisation techniques. By implication, they also provide guidance on when the standard of Functional Legal Anonymisation may be deemed likely to be achieved (i.e.

[478] Compare, e.g. the language of the DPA (section 1(1)). Although it does not include reference to the Means Test, it does *not* suggest that the results of anonymisation should be without residual risk of re-identification. This is because the converse of its definition of personal data is: data which *does not* relate to a living individual who can be identified – (a) from those data, or (b) from those data and other information which is in the possession of, *or is likely to come into the possession of*, the data controller. In other words, it is possible that a data subject could still be re-identifiable through the modified data in combination with additional information *not likely to come into the possession of the data controller.* Notwithstanding, like the DPD, the DPA does not provide any practical assistance to help organisations determine whether anonymised data they share is likely to result in the re-identification of an individual (legally).

[479] Recital 26 DPD states: *"whereas codes of conduct …may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible*."

taking into account all means likely reasonably that could be used either by the controller or by any other person to re-identify individuals regarding data made subject to anonymisation processes).

5.3.1.1 The WP

First, we return to WP136, before looking at key guidance issued by the WP in WP216 (its 2014 Opinion on Anonymisation Techniques). It is useful, in this respect, to consider how the WP's views appear to have evolved over the seven-year time-span between these opinions.

5.3.1.1.1 WP 136 (2007)

WP136 describes pseudonymisation as *"the process of disguising identities…to be able to able to collect additional data relating to the same individual without having to know his identity"*.[480] It also comments that pseudonymised data are information about individuals that are indirectly identifiable, and thus personal data, where the pseudonymisation is retraceable.[481] Said otherwise, WP136 acknowledges that the legal status of data (as personal data or non-personal data) that has undergone pseudonymisation depends on related circumstances.[482] For example, in a key-coding data scenario,[483] the WP highlights the types of re-identification risk factors for assessment:

> If the codes used are unique for each specific person, the risk of identification occurs whenever it is possible to get access to the key used for the encryption. Therefore the

[480] WP136, p.18.

[481] Ibid, p.18: "*[p]seudonymisation can be done in a retraceable way by using correspondence lists for identities and their pseudonyms or by using two-way cryptography algorythms for pseudonymisation… retraceably pseudonymised data may be considered as information on individuals which are indirectly identifiable. Indeed, using a pseudonym means that it is possible to backtrack to the individual, so that the individual's identity can be discovered*…" The WP goes on to say on the same page that where a pseudonym is used, this identity re-identification would only happen under predefined circumstances and in that case, "*although data protection rules apply, the risks at stake for the individuals with regard to the processing of such indirectly identifiable information will most often be low, so that the application of these rules will justifiably be more flexible than if information on directly identifiable individuals were processed*". The WP contrasts retraceable pseudonymisation on the same page with a process of *"disguising identities [that] can also be done in a way that no reidentification is possible, e.g. by one-way cryptography, which creates in general anonymised data".* The latter phrase – "general anonymised data" is particularly ambiguous in terms of the legal implications it intended to suggest here.

[482] In fact, the WP considers this not just to be relevant to data from which direct identifiers have been deleted altogether, but also pertinent to statistical information where, even though information is aggregated, the size of the original group must be taken into account. If that size is small, it says, identification may still be possible through combining the aggregated information with other information. See ibid, p.18.

[483] This is described by the WP as where unique codes replace common identifiers of individuals (like name, date of birth, address) in data and *"the key making the correspondence between the code and the common identifiers …is kept separately"* (ibid, p.18). If codes used are not unique to individuals, however, WP136 seems to consider that the risk of identification might even be eliminated, and the data rendered no longer personal. Such measures may involve aggregation of the data; in the examples given in WP136, the same code number is use for all individuals in the same town or for all records for the same year. See also p.19: "*[i]f, on the contrary, the codes are not unique, but the same code number (e.g. "123") is used to designate individuals in different towns, and for data from different years (only distinguishing a particular individual within a year and within the sample in the same city), the controller or a third party could only identify a specific individual if they knew to what year and to what town the data refer. If this additional information has disappeared, and it is not likely reasonably to be retrieved, it could be considered that the information does not refer to identifiable individuals and would not be subject to the data protection rules"*.

> risks of an external hack, the likelihood that someone within the sender's organization - despite his professional secrecy - would provide the key and the feasibility of indirect identification are factors to be taken into account to determine whether the persons can be identified taking into account all the means likely reasonably to be used by the controller or any other person…[484]

Notwithstanding, WP136 recognises that even this type of retraceably pseudonymised data may be considered non-personal data in third-party hands in circumstances where appropriate measures have been taken to prevent re-identification by the third party.[485] Thus, WP136 adopts a relativist perspective on the notion of (re-)identifiability (the same data may be personal or non-personal depending on from whose eyes re-identifiability is being assessed). Hence, for the WP in 2007, the status of information relating to persons that has had pseudonymisation procedures applied to it is unclear without further analysis in each case and depending on the perspective of the one being considered. Whether information amounts to personal data depends on the circumstances, and consideration of all means likely reasonably used to identify individuals, including the measures used to prevent re-identification. Furthermore, WP136 notes that de-anonymisation techniques will improve over time so reappraisal of the type and level of re-identification risks by the relevant data controller is necessary.[486]

5.3.1.1.2 WP216 (2014)

WP216 covers a range of legal, policy, and technical issues surrounding personal data anonymisation. It is not a systematic manual on how to go about anonymisation. Rather, it outlines a good practice framework to enable data controllers to make better decisions about carrying out anonymisation, particularly in an information-sharing context.[487]

WP216 describes popular anonymisation techniques to assist data controllers with designing anonymisation programmes and highlights each of their respective strengths/weaknesses. There

---

[484] Ibid, p.19.

[485] Ibid, p.20: *"[t]his does not mean, though, that any other data controller processing the same set of coded data would be processing personal data, if within the specific scheme in which those other controllers are operating re-identification is explicitly excluded and appropriate technical measures have been taken in this respect".*

[486] Ibid p.15: "*[o]n the other hand, this test is a dynamic one and should consider the state of the art in technology at the time of the processing and the possibilities for development during the period for which the data will be processed. Identification may not be possible today with all the means likely reasonably to be used today. If the data are intended to be stored for one month, identification may not be anticipated to be possible during the "lifetime" of the information, and they should not be considered as personal data. However, it they are intended to be kept for 10 years, the controller should consider the possibility of identification that may occur also in the ninth year of their lifetime, and which may make them personal data at that moment. The system should be able to adapt to these developments as they happen, and to incorporate then the appropriate technical and organisational measures in due course".*

[487] WP216 starts (p. 5) by acknowledging the importance of data sharing to individuals and society and the significant role of anonymisation techniques in this respect, including the potential benefits derived from using them as part of open data strategies in protecting privacy rights.

are three techniques described as those using: randomisation (noise addition, permutation, or differential privacy); generalisation (aggregation and K-anonymity, L-diversity/T-closeness); and, pseudonymisation (encryption with secret key, hash function, keyed hash-function with stored key, deterministic encryption or keyed-hash function with key deletion, or tokenisation).

However, there is a duality of identificatory-approach interpretational models evident in WP216. On the one hand, the WP says it interprets "anonymised data" as meaning "*anonymous data that previously referred to an identifiable person, but* ***where that identification is no longer possible***" (emphasis added).[488] On the other hand, in other parts of WP216, the WP retreats from the irreversibility ideal in favour of a pragmatic attitude lauding the Functional Legal Anonymisation standard (i.e. founded upon satisfying the Means Test) as the one to achieve, requiring the consideration of contextual elements.[489]

WP216 also mentions an aid to assessing re-identification risk by postulating a possible adversary trying to re-identify a data subject from transformed data for their own purposes. Thus, any steps taken by data controllers aimed at preventing data subjects being re-identified should encompass consideration of "all" the means "reasonably likely" to be used by a hypothetical adversary in the circumstances. A key component of the definitional focus should also be on measures implemented to mitigate the likelihood of a re-identification outcome in the future.[490] Against this backdrop, the WP develops its analysis on why it believes that there is *no* guarantee that pseudonymising personal data will satisfy the Functional Legal Anonymisation standard.

First, in recommending improvements to the effectiveness of each different type of anonymisation method described, the WP states that each "*fails to meet with certainty the criteria of anonymisation*". Such WP216 "criteria" view the robustness of each technique to mitigate the likelihood of further re-identification of a particular individual from data ("*performed by the most*

---

[488] Ibid, at p.8. WP216 also alludes to *"irreversibility of the alteration undergone by personal data to enable direct or indirect identification"* as key to definitions of anonymisation in international standards (p.6); and, in describing the anonymisation process, states (p.5) *"[a]n important factor is that the processing must be irreversible*".

[489] Ibid, pp. 8-9: "*it can be argued that data controllers should focus on the concrete means that would be necessary to reverse the anonymisation technique, notably regarding the cost and the know-how needed to implement those means and the assessment of their likelihood and severity. For instance, they should balance their anonymisation effort and costs (in terms of both time and resources required) against the increasing low-cost availability of technical means to identify individuals in datasets, the increasing public availability of other datasets…"*

[490] For example (ibid, p.8) relevant considerations include any informational disclosure control mechanisms put in place by the data controller to restrict access to the data. Falling short of advocating the existence of a technical panacea to the possibility of eliminating re-identification risk, the WP advocates using combinations of the best techniques currently available to reduce re-identification risk, planning anonymisation on a case-by-case basis, as well as factoring in other WP recommendations. For example, data controllers are advised not to treat anonymisation as a one-off exercise because re-identification risks can change over time (e.g. with the release of future data). Rather, regular risk-assessment should continue in the light of the residual risk of re-identification. This suggests that for the WP in WP216 the best approach for a data controller to take in planning future anonymisation processes may also be dependent on the context of the processing for which transformed data are to be used.

likely and reasonable means the data controller and any third party may employ"[491]) based on three different types of re-identification risks from data:[492]

- **The possibility of singling-out an individual in a dataset(s),** "*which corresponds to the possibility to isolate some or all records which identify an individual in the dataset*";[493]
- **The possibility of linking records relating to an individual ('linkability') in a dataset(s)**, "*which is the ability to link, at least, two records concerning the same data subject or a group of data subjects (either in the same database or in two different databases*";
- **The possibility of inferring information about an individual ('inference') in a dataset(s)**, *"which is the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes*".

Second, the WP comments, "*pseudonymised data … continue to allow an individual data subject to be singled out and linkable across different data sets*".[494] It adds, "*pseudonymisation is not a method of anonymisation"*, but "*merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure*".[495] Said otherwise, the WP confirms that, in its opinion, "*[p]seudonymity is* ***likely to allow for*** *identifiability,* ***and therefore stays inside the scope of the legal regime of data protection***" (emphasis added).[496] The latter WP216 statement, especially, suggests that data that has been pseudonymised should *always* be deemed personal data by way of a *presumption* of a certain degree of likeliness generally (in contrast to the WP's more equivocal position on this issue in WP136).

Another striking distinction between WP136 and WP216 is the re-identifiability perspective taken. In WP216, the WP appears to abandon its earlier relativist approach for a more objectivist position, or potentially even absolutist stance.[497] Specifically, WP216 suggests that if data controllers retain raw personal data at source, they also retain the ability to attribute it to the relevant individuals;

---

[491] Ibid, p.12.
[492] Ibid, pp.9, 11- 12.
[493] WP216, therefore, marks some change in the WP's position regarding the scope of risks entailed by processing pseudonymised data from its position in 2007. That is because it suggests that even irreversibly (untraceably) pseudonymised data can allow an individual data subject to be singled out and carries re-identification risk (contrast the suggestion from the quote cited at fn.488 above).
[494] Ibid, p.10.
[495] Ibid, p.3 (executive summary).
[496] Contrast WP136, p. 20: "*[t]his does not mean, though, that any other data controller processing the same set of coded data would be processing personal data, if within the specific scheme in which those other controllers are operating re-identification is explicitly excluded and appropriate technical measures have been taken in this respect."* In general, however, under WP136 and WP216, it is suggested by the WP that the application of pseudonymisation to personal data is not sufficient, on its own, to eradicate re-identification risk to a sufficient level that the data becomes non-personal data legally.
[497] As a reminder from Chapter 3, in accordance with an absolutist stance, as long as one person holds identifying data then raw data is personal data because someone (anyone at all) can use it to identify the data subject.

and, for that reason, its modified form remains personal data.[498] As explained by Knight/Stalla-Bourdillon, this statement has several implications:

> By affirming such statements, and despite what is says elsewhere in the same Opinion, the Art. 29 WP appears to be rejecting the very consequences of a risk-based approach. This is because, if it is possible to isolate the raw datasets from the transformed datasets and put in place security measures, including technical and organisational measures, as well as legal obligations (essentially contractual obligations), so that the subsequent recipient of the transformed dataset will never have access to the raw dataset, the transformed dataset should be deemed as comprising data rendered anonymous at the very least in the hands of the subsequent recipient of the dataset.[499]

---

[498] WP216, p.9: *"it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data*". The only exception, the WP says, is where "*the data controller would aggregate the data to a level where the individual events are no longer identifiable*". The WP gives the following example*: "if an organisation collects data on individual travel movements, the individual travel patterns at event level* ***would still qualify as personal data for any party, as long as the data controller (or any other party) still has access to the original raw data,*** *even if direct identifiers have been removed from the set provided to third parties. But if the data controller would delete the raw data, and only provide aggregate statistics to third parties on a high level, such as 'on Mondays on trajectory X there are 160% more passengers than on Tuesdays', that would qualify as anonymous data"* (emphasis added).

[499] Knight & Stalla-Bourdillon (2016, p.299). This quote goes on, "*[t]his is all the more warranted…because where the transformed dataset is still considered to be personal data in the hands of the subsequent recipient of the dataset, complying with the entire gamut of data protection obligation is likely to have a chilling effect on data sharing given the complexity of distilling the new compliance obligations that will be required to be followed by organisations under the GDPR. This could have, in particular, a significant deterrent effect on the carrying out of beneficial longitudinal research studies that rely upon the reuse of personal data once it has undergone anonymisation methods, such as in the health or education sectors."* See also El Emam & Álvarez (2014, p.86): "*the Opinion is moving in the direction of prohibiting longitudinal data. In order to create longitudinal trails of individuals, such as multiple visits to a clinic or multiple treatments during a hospital stay, it is absolutely critical to be able to link the records that belong to the same individual. If anonymization techniques break linkability within the same database then many large data sets would be quite useless".* To more fully explain the reason for this chilling effect in a longitudinal context, consideration should be given to how the personal data anonymised was obtained originally and the feasibility of compliance with data protection rules in respect of secondary usage retrospectively. See comments by the ICO in its Anonymisation Code of Practice discussed in the next sub-section (2012, p.29): "*[i]f, for example, the data was collected as part of a survey and individuals were told that it would be used for research purposes then clearly there will be no barrier to using the data for that purpose. In some cases individuals may have been given an assurance that personal data about them would only be used for a particular purpose, eg to despatch the goods they have ordered. Assurances of this nature should be respected, but very specific purpose limitation of this type is rare. A more common scenario is for an organisation to have a collection of personal data obtained for a particular purpose or set of purposes, eg to administer individuals library services. In cases like this individuals may never have been notified as to whether their data will or will not be anonymised for use in research purposes, for example. Organisations should address this in their privacy policies or by other means. … even if individuals have no choice over the provision of their personal data, this does not mean that they have the right to stop the organisation anonymising it provided the processing of their personal data in order to anonymise is not likely to cause unwarranted damage or distress. Of course the processing of personal data must also comply with the data protection principles, meaning it must be fair, for example. It is also generally unfeasible to see data return (ie recalling data or removing it from a website) as a safeguard given the difficulty, or impossibility, of securing the deletion or removal of data once it has been published".* Thus, data controllers could become subject to data protection law in respect of the secondary processing of anonymised data by recipients even if appropriate measures have been taken to mitigate re-identification risk in practice taking accounts of the 'means likely reasonably' available to third parties. The recipients may also take on their own data protection responsibilities as joint controllers, or processors of the data. To note, in engaging a processor to process personal data for it, a controller must choose one providing guarantees of a sufficient nature in respect of the technical security measures and organisational measures governing the processing to be carried out, and the controller must ensure compliance with those measures. Under the GDPR (Article 28), processors will also take on independent compliance obligations.

At the least, WP216 promotes uncertainty as to which interpretational model of the non-personal data concept is the correct one under the identificatory-approach. Moreover, this uncertainty is compounded by the WP's implication of new features into the Means Test not mentioned elsewhere, e.g. by stating that it should be assessed against a standard of whether identification of the data subject has become "*reasonably impossible*".[500]

5.3.1.2 The ICO

5.3.1.2.1 Anonymisation code of practice (2012)

Two years before WP216, the ICO published its Anonymisation Code of Practice (hereafter, **the Code**). The Code was also intended to provide a framework for data controllers to use when considering an anonymisation strategy to bring "*a greater consistency of approach and to show what [the ICO] expect of organisations*" under the DPA, rather than intended as a step-by-step manual of how-to-do anonymisation.[501]

The Code remarks that the effective anonymisation of personal data is both "*possible*"[502] and "*desirable*"[503] under a re-identification risk-based approach, and sets forth recommendations for determining when data should be considered non-personal data under the DPA. Specifically, the ICO states that anonymisation techniques that depend upon removing individual identifiers from person-specific data carry higher risks, but not insurmountable ones.

In describing ways of assessing re-identification risk, the Code places greatest emphasis on evaluating the likely availability of other information to a third party (by which they might be able to re-identify an individual alongside the data). The ICO describes this task as "*extremely problematic"* as *"more potentially 'match-able' information becomes publicly available*";[504] moreover, re-identification risk through data linkage *"is essentially unpredictable because it can never be assessed with certainty what data is already available or what data may be released in the future"*.[505] The ICO also refers to many borderline cases requiring "*careful judgement…based on the*

---

[500] WP216, p.8 (in full): "*the "means … reasonably to be used" test is suggested by the Directive as a criterion to be applied in order to assess whether the anonymisation process is sufficiently robust, i.e. whether identification has become "reasonably" impossible*".

[501] ICO (2012, New anonymisation code sets out how to manage privacy risks and maintain transparency, [online]).

[502] The Code, p.7, despite the fact that the ICO suggests that identification from data can be achieved merely by being "*able to establish a reliable connection between particular data and a known individual".*

[503] Ibid, with the ICO also remarking that it, "*can help society to make rich data resources available whilst protecting individuals' privacy*". The ICO also comments on the fact that the effective anonymisation of personal data is of particular relevance now, "*given the increased amount of information being made publicly available through Open Data initiatives and through individuals posting their own personal data online*".

[504] The Code, p.18. For example, *"[t]he 'other information' needed to perform re-identification could be information available to certain organisations, to certain members of the public or that is available to everyone because it has been published on the internet".*

[505] The Code, p.18.

circumstances", where it will be "*difficult or even impossible to determine whether it is likely that re-identification will take place".*[506] Such statements also link to discussion around the objective evidential basis that should underpin re-identification determinations under the DPA.[507]

The Code recommends adopting a 'motivated intruder' test to help data controllers assess the level of re-identification risk that arises regarding the potential for a successful third party attempt to re-identify data subjects.[508] This test refers, says the ICO, to the postulation of a hypothetical entity "*who starts without any prior knowledge but who wishes to identify the individual from whose personal data the anonymised data has been derived*".[509] The ICO extends its analysis by encouraging consideration of whether such a hypothetical attacker would be able to identify an individual, assuming they have certain motivation, means, and skills.[510] The Code further highlights the importance of secondary factors in informing an organisation's approach to data disclosure and security, such as the type of data at issue but also who it is about - and the organisation holding the data (e.g. because an intruder/campaigner wants to show that anonymisation undertaken by such

---

[506] The Code, p.18 and p.20. See also p.21: "*[w]hat is the risk of re-identification? In some cases the risk of anonymised data being combined with other data to result in personal data being created will be high….However, in some circumstances it can be difficult to establish the risk of re-identification, particularly where complex statistical methods might be used to match various pieces of anonymised data".*

[507] The Code, p.26: "*Information, established fact and knowledge When considering re-identification risk, it is useful to draw a distinction between recorded information, established fact and personal knowledge: Established fact might be that Mr B Stevens lives at 46 Sandwich Avenue, Stevenham. This could have been established by looking at an up-to-date copy of the electoral register. Personal knowledge might be that I know Mr B Stevens is currently in hospital, because my neighbour Mr Stevens wife told me so. The starting point for assessing re-identification risk should be recorded information and established fact. It is easier to establish that particular recorded information is available, than to establish that an individual or group of individuals - has the knowledge necessary to allow re-identification. However, there is no doubt that non-recorded personal knowledge, in combination with anonymised data, can lead to identification. It can be harder though to substantiate or argue convincingly. There must be a plausible and reasonable basis for non-recorded personal knowledge to be considered to present a significant re-identification risk. Identification and the educated guess Data protection law is concerned with information that identifies an individual. This implies a degree of certainty that information is about one person and not another. Identification involves more than making an educated guess that information is about someone; the guess could be wrong.* ***The possibility of making an educated guess about an individual's identity may present a privacy risk but not a data protection one because no personal data has been disclosed to the guesser.*** *Even where a guess based on anonymised data turns out to be correct, this does not mean that a disclosure of personal data has taken place"* (emphasis added). See also the TGN where the ICO states, in relation to the issue of identifiability (p.8), "*…the fact that there is a very slight hypothetical possibility that someone might be able to reconstruct the data in such a way that the data subject is identified is not sufficient to make the individual identifiable for the purposes of the Directive.* ***The person processing the data must consider all the factors at stake***" (bold emphasis added).

[508] The Code, pp.22-24. To note, the 'motivated intruder' test is also mentioned by the WP in an anonymisation context in WP203 (p.31): "*[i]n practice, there is a very significant grey area, where a data controller may believe a dataset is anonymised, but a motivated third party will still be able to identify at least some of the individuals from the information released*".

[509] The Code, p.22.

[510] These include access to resources (such as libraries and the Internet), in addition to the ability to employ investigative techniques (such as "*making enquiries of people who may have additional knowledge of the identity of the data subject or advertising for anyone with information to come forward*"). Ibid, p.22-23. Despite being attributed with reasonable competence, the attacker is not presumed – says the ICO - to have any "*specialist knowledge such as computer hacking skills, or to have access to specialist equipment or to resort to criminality such as burglary, to gain access to data that is kept securely*".

an organisation had previously been unsafe, e.g. Netflix/AOL incidents mentioned in Chapter 1) - which could motivate an intruder to attempt re-identification.[511]

Hence, the ICO takes a more flexible approach than the WP takes two-years later in analysing the nature of re-identification risk. By not referencing a 'linkability' element to its evaluation of the standard for effective anonymisation, the Code seems to set the regulatory bar lower than WP216.[512] Furthermore, according to a relativistic perspective on re-identifiability,[513] the Code's bottom line is that effective anonymisation through pseudonymisation (defined by it as "*[t]he process of distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity*") is possible.[514] The key, for the ICO, lies in data controllers mitigating the likelihood of re-identification risk associated with a particular processing activity sufficiently from

[511] Ibid, p.23: "*[c]learly, some sorts of data will be more attractive to a 'motivated intruder' than others. Obvious sources of attraction to an intruder might include: • finding out personal data about someone else, for nefarious personal reasons or financial gain; • the possibility of causing mischief by embarrassing others; revealing newsworthy information about public figures; • political or activistic purposes, eg as part of a campaign against a particular organisation or person; or • curiosity, eg a local person's desire to find out who has been involved in an incident shown on a crime map".*

[512] In other words, when WP216 adopts a Functional Legal Anonymisation standard, that standard appears to be interpreted more strictly than in the Code. Although to note, the ICO does refer to 'linkage' once in the Code as a side-consideration, see p.24: "*[m]otivated intruder risk: some issues to consider - What is the risk of jigsaw attack, ie piecing different bits of information together to create a more complete picture of someone? Does the information have the characteristics needed to* ***facilitate data linkage - eg is the same code number used to refer to the same individual in different datasets****? What other* ***linkable information*** *is available publicly or easily?"* (emphasis added).

[513] Ibid, p.21: "e*ven though pseudonymised data does not identify an individual,* ***in the hands of those who do not have access to the 'key', the possibility of linking several anonymised datasets to the same individual can be a precursor to identification***" (emphasis added). To note, the ICO appears to support a re-identification risk-assessment perspective that is not absolutely relativist or objectivist, at least in the sense of considering such risk not just from the perspective of data in the hands of one person entirely (whoever that person may be), but 'in the round'. This phrase means that all organisations disclosing data should assess whether anyone else could identify any individual from the data being released, either by itself or in combination with other available information. For example, the ICO links assessing identifiability from data 'in the round' with the 'motivated intruder test' (p.19): "*different members of the public may have different degrees of access to the other information needed for re-identification to take place. However, a motivated intruder test can go some way towards addressing this problem. It is good practice to try to look at identification in the round, ie all organisations disclosing anonymised data should assess whether any organisation or member of the public could identify any individual from the data being released either in itself or in combination with other available information. The risk involved will vary according to the local data environment and particularly who has access to information. This means that anonymised data disclosed within a secure local environment, eg when disclosed to a particular research organisation, could remain anonymous even though if published, the likelihood of re-identification would mean that the anonymised data would become personal data".*

[514] Ibid, p.21. Despite this, the ICO also acknowledges in the Code (p. 20) that, *"[t]here will clearly be borderline cases where, in reality, it will be difficult, or even impossible, to determine whether it is likely that re-identification will take place".*

the perspective of the recipient.[515] Yet, this is a dynamic process requiring the carrying out of risk-assessments, and managing risks identified, periodically.[516]

5.3.1.2.2 Big data and data protection (2014, rev.2017)

In its 2014 paper exploring the implications of Big Data for data protection, the ICO again advises organisations on how to carry out robust re-identification risk-assessments. While admitting that it may be impossible to establish with certainty that re-identification risk does not exist in any particular case, it says "*that does not mean anonymisation is impossible or that is not an effective tool*".[517] It then immediately endorses the DPD Means Test:

> If personal data is fully anonymised, it is no longer personal data. In this context, anonymised means that it is not possible to identify an individual from the data itself or from that data in combination with other data, taking account of all the means that are reasonably likely to be used to identify them.[518]

Strikingly, in comparison with the Code,[519] the paper introduces some evolution in the ICO's thinking regarding the level of re-identification risk that it deems acceptable to meet the Functional Legal Anonymisation standard:

---

[515] The Code notes (p.16) that: *"[t]he High Court in [R (on the application of the Department of Health) v Information Commissioner [2011] EWHC 1430 (Admin)] stated that the risk of identification must be greater than remote and reasonably likely for information to be classed as personal data under the DPA."* Indeed, such is the assuredness of the ICO in the Code that effective anonymisation through pseudonymisation is possible, it says it is "*confident that adopting the techniques and procedures recommended in this code will guard against re-identification"* (p.27). These recommendations, notably, also include compliance with certain substantive obligations under the DPA itself (e.g. ibid, p. 19): "*[d]espite all the uncertainty, re-identification risk can certainly be mitigated by ensuring that only the anonymised data necessary for a particular purpose is released. The fact that data has been anonymised does not mean that data minimisation techniques are not still relevant".*

[516] Ibid, p. 25: "*[d]ata controllers must be aware of the risk of re-identification and that this risk can change over time, eg powerful data analysis techniques that were once rare are now common-place. However, if anonymisation is carried out effectively in the present this is likely to protect personal data from future re-identification. A realistic assessment of the risk of re-identification occurring in the future should be made, meaning that organisations should not assume that data that is anonymous now will necessarily become re-identifiable in the future. However, organisations should carry out a periodic review of their policy on the release of data and of the techniques used to anonymise it, based on current and foreseeable future threats*". In other words, the ICO clearly supports a risk-management approach, as well as periodical review of new data released and techniques for re-identification. From a pragmatic viewpoint, for example, the Code says that it is good practice, when releasing anonymised data to try to assess (p.21): "*the likelihood of individuals having and using the prior knowledge necessary to facilitate re-identification. It is accepted that this will be difficult to conduct on a record by record basis for large datasets or collections of information. It will often be acceptable to make a more general assessment of the risk of prior knowledge leading to identification, for at least some individuals recorded in the information and then make a global decision about the information; the chances that those who might be able to re-identify are likely to seek out or come across the relevant data*".

[517] ICO (2014, #1, para.42, p.12). In the 2017 version of this document (ICO, 2017 #2, para.134, p.60), this statement is rephrased to "*that does not make anonymisation impossible or ineffective*". Thus, again, there is no suggestion that the ICO subscribes to a zero-risk standard of non-personal data as a concept.

[518] ICO (2014, #1, para.40, p.11). In the 2017 version of this document (ICO, 2017 #2, para.130, p.58), this statement is unaltered.

[519] Albeit that there is also an equivocal statement by the ICO suggesting that it might now prefer a more objective (but not absolute) perspective on (re-)identifiability in ICO (2014, #1) and copied in ICO (2017, #2, para.136, p.60): *"an organisation may hold a dataset containing personal data in one data store, and produce an anonymised version of it to be used for analytics in a separate area.* ***Whether it remains personal data*** *will depend on whether the anonymisation*

> The issue is not about eliminating the risk of re-identification altogether, but whether it can be mitigated **so it is no longer significant**. Organisations should focus on mitigating the risks **to the point where the chance of re-identification is extremely remote**.[520] (emphasis added)

Said otherwise, the ICO adopts an 'extremely-remote' risk standard of re-identification likelihood to be met for effective anonymisation (without making it clear if a 'significant risk' of re-identification should be directly equated with a chance of re-identification that is more than (extremely) remote, although that is the conclusion that can be drawn from its statements), while also emphasising the importance of re-identification risk-assessment and risk-mitigation steps being carried out so that organisations using anonymised data can demonstrate that they have carried out "*robust*" assessments, and adopted solutions "*proportionate to the risk*".[521] In this respect, the ICO makes comparisons between the process of anonymisation, and the use of privacy-risk mitigating tools (such as PIAs), as both requiring consideration of the use of anonymisation techniques – i.e. they share a common objective in intending to protect privacy by mitigating re-identification risk *and* the risk of data misuse.[522]

5.3.1.3 Judicial guidance

There are no CJEU cases on anonymisation and data protection law. By comparison, there is useful UK case-law interpreting the DPA in respect of re-identification risk from statistically-aggregated data. However, as such case-law does not involve determining the meaning of anonymisation in respect of the status of modified data retaining individualised data points, a summary of the main case law body (and its critical relevance to addressing sub-research question 1) is described separately in Appendix 3 below.[523] In distinction, discussed next is an English case about data enabling the singling-out of individuals worthy of detailed consideration because of the unique issues it raises.

"keys" and other relevant data that enable identification are retained by the organisation. Even if the data remains personal data this is still a relevant safeguard to consider in order to enable processing to comply with the data protection principles" (emphasis added)).

[520] In ICO (2014, #1) and copied in ICO (2017, #2, para.134, p.60).

[521] In ICO (2014, #1) and copied in ICO (2017, #2, para.135, p.60).

[522] In ICO (2014 #1) and copied in ICO (2017#2, para.139, p. 61): "*[a]nonymisation should not be seen merely as a means of reducing a regulatory burden by taking the processing outside the DPA. It is a means of mitigating the risk of inadvertent disclosure or loss of personal data, and so is a tool that assists big data analytics and helps the organisation to carry on its research or develop its products and services. It also enables the organisation to give an assurance to the people whose data it collected that it is not using data that identifies them for its big data analytics. This is part of the process of building trust which is key to taking big data forward".*

[523] See also the useful case-law overview described by Aldhouse (2014).

In 2014, Queen Mary University of London (QMUL) received a freedom of information request for clinical trial participant data relating to chronic fatigue syndrome and its treatment in an "anonymised" state; notwithstanding, also requested was the formatting of such data such that each row contained "*values from the same [trial] participant*".[524] QMUL refused the request on the basis, inter alia, that the data constituted personal data under the DPA precluding its disclosure under section 40(2) FOIA.[525] Specifically, QMUL asserted that *"[t]he data [to be released] is pseudonymised, not anonymised, and therefore is likely to constitute personal data"* because it has *"individual-level granularity"* giving rise to a *"relatively high risk" of re-identification".*[526]

[524] In background, QMUL collected medical baseline and treatment results from 640 participants in a clinical chronic fatigue syndrome (myalgic encephalomyelitis (ME)) trial from 2005-2010. The FOIA request was for a *"selection of baseline and 52-week follow up data on all 640 individual PACE Trial participants for which the data exists, in a spreadsheet or equivalent file with separate columns for each variable."* The requester then added, *"I am requesting only 'anonymised' data, I am not requesting any information which can identify individual participants (not even the participant ID numbers if those are deemed to be inappropriate to include, so long as each individual row only contains values from the same participant)."* In particular, each row of the data requested contained 14 columns, with the first column containing the personal pin number for each participant, and the remaining columns containing numbers that represent the outcomes of various tests related to the participant. Much of these test 'results' were calculated from participant-completed questionnaires (that did not contain identifiers such as location, gender or ethnicity), questions answered verbally, or physical tests (e.g. walking tests). Participants were assured strict confidentiality of the data collected from them during the trial. On their consent forms, however, they were also informed, and agreed to, the possibility of such data being shared with non-QMUL scientists with whom QMUL would sign confidentiality agreements (by way of normal research collaboration procedures). The trial results were first published in 2011, and a follow-up study concluded in 2012, after which analysis of the data continued along with publication of academic papers discussing findings.

[525] Per references in Chapter 3, Section 40(2) FOIA states: *"(2) Any information to which a request for information relates is also exempt information if—(a) it constitutes personal data which do not fall within subsection (1), and (b) either the first or the second condition below is satisfied. (3)The first condition is—(a) in a case where the information falls within any of paragraphs (a) to (d) of the definition of "data" in section 1(1) of the Data Protection Act 1998, that the disclosure of the information to a member of the public otherwise than under this Act would contravene—(i) any of the data protection principles, or (ii) section 10 of that Act (right to prevent processing likely to cause damage or distress), and (b) in any other case, that the disclosure of the information to a member of the public otherwise than under this Act would contravene any of the data protection principles if the exemptions in section 33A(1) of the Data Protection Act 1998 (which relate to manual data held by public authorities) were disregarded.(4)The second condition is that by virtue of any provision of Part IV of the Data Protection Act 1998 the information is exempt from section 7(1)(c) of that Act (data subject's right of access to personal data)."*

[526] Queen Mary University of London v The Information Commissioner and another [2016] EA/2015/0269 (12 August 2016) (Information Tribunal), p.10. QMUL took the view (quoted in the IC's 2015 FOIA decision FS50565190, at, p.14) that the data requested was *"individual level data which by its very nature "are much more likely to reveal an individual's identity than aggregate data", a risk that increases as the number of data items increases and with the number of individual-level records (see HSCIC Anonymisation Standard (2013))".* (In background, to note as part of its arguments, QMUL also referred to a separate 2013 FOIA decision (FS50514995) in which the IC *had* upheld the application of section 40(2) FOIA to a request for primary outcome measure scores from the same clinical trial. The IC noted in that decision that the request considered there asked that each set of information should be similarly ordered by e.g. recruitment date or random participant number assignment. QMUL argued that identification would be possible if this type of information became available in the public domain in combination with the data requested. The IC accepted this argument on the facts in terms of the possibility of combining information with randomised dates made public).

Returning to the particular litigation under discussion, in the IC's decision FS50565190 (p.10), QMUL *also* argued that*: "[e]ven if the disclosure risk for third parties to be able to identify individuals were low, the possibility that an individual who took part in the trial could identify themselves is much higher, especially since much of these health data were self-rated. It is not possible therefore to render the data completely anonymous given this, the quantity of participants and the combination of data fields requested."* Furthermore, if some can self-identify, related others may also be able to do so. Finally, QMUL argued that disclosing data to the FOIA requester would be against the express assurances given to participants regarding confidentiality.

QMUL's refusal was appealed to the Information Commissioner (IC), who held that the requested information was anonymised effectively and *not* personal data.[527] On further appeal, a majority in the First-Tier (Information Rights) Tribunal (hereafter 'FTT') upheld the IC's decision on this point.[528] The FTT's reasoning in its decision (hereafter 'FTT judgement') can be summarised as follows.

- **(Re-)identification how? -** Even though the data retained participant-level data points (one-row-per-one-participant), it contained no direct or fixed identifiers. The FTT concludes that Functional Legal Anonymisation was still possible under the DPA. Said otherwise, the ability to single-out data subjects from the modified data requested was not enough without consideration of whether the Means Test was also met.

- **(Re-)identification from whose perspective?** - QMUL posited three possible routes to identifiability from the requested data: (1) participants could self-identify; (2) those with prior knowledge of participants (e.g. friends, family, or medical practitioners) could identify them;[529] or (3) motivated intruders (such as campaigners or journalists) could identify participants by linking the released data to other information previously released. The FTT approves the use of the ICO-recommended motivated intruder test, in assessing the evidence of the risk of a hypothetical-type of individual (matching certain specifications) being able to identify participants. However, its endorsement of the Means Test is caveated by the requirement that an evidential basis be proffered regarding the conceivability in practice of the identificatory-means hypothesised.[530] This view appears at odds with

[527] IC 2015 FOIA decision FS50565190, p.12 referring to arguments used by the ICO as part of its decision: *"anonymization is plainly capable of rendering those individuals non-identifiable for two reasons: (i) The pool of participants is large, and the incidence of chronic fatigue/ME in the general UK population is around 1 %, therefore the class of potential trial participants 'vastly exceeds' the number of actual participants, rendering identification realistically impossible; (ii) The information is not directly linked to the individuals, as it comprises wide-ranging scores derived from participants' self-reporting."*

[528] Queen Mary University of London v. The Information Commissioner and another [2016] EA/2015/0269 (12 August 2016) (Information Tribunal).

[529] In terms of the first two routes to identification, the IC had discarded them for the reason that – in its opinion, for an individual to be identifiable – it must be reasonably likely that they or another person can identify them from that information and other information that may be available to them. In other words, identification is more than making an educated guess about identity. See, e.g., IC decision FS50565190, p.18: "*[w]hilst the Commissioner acknowledges the possibility that some individuals may be able to identify themselves in the withheld information, he does not believe that this is sufficient for those individuals to be identifiable for the purposes of section 40(2)".* Hence, the IC argued that identification for the purposes of the DPA must be able to be made by a third party. In any event, the IC added, there is no evidence that any participant actually retained their exact scores and information from the trial. Regarding the second route to identification, it is deemed unlikely that close friends and family would be motivated intruders (i.e. realistically be able to identify participants), and even less likely that they will have obtained the information in the first place, let alone retained it after the passage of years between the trial and the request, and certainly highly unlikely with the necessary precision to facilitate future identification.

[530] To this end, the FTT reiterates that the test requires consideration of whether an individual is likely reasonably to have the means and skill to identify any participants, in addition to whether he/she is likely reasonably to use those skills for that purpose. Compare, e.g. the IC's view (in IC decision FS50565190, p.18) that QMUL did not provide, "*any evidence as to how the motivated intruder might be able to actually identify participants from the trial from the information contained in the requested information and other information that may be available to such an individual";* nor did it indicate *"what means are reasonably likely to be available to the motivated intruder to facilitate re-*

comments in the Code implying that controllers need not be expected to know the likelihood that potential re-identifiers (those that have and use prior knowledge necessary to facilitate re-identification) will actually come across relevant additional information (nor exactly what relevant data is 'out there' presently).[531]

- **(Re-)identification risk to what degree? –** The FTT confirms that anonymisation does not need to be completely risk-free to be effective, but the risk of (re-)identification from data disclosure (including by a motivated intruder) must be stronger (more likely) than 'remote' on the evidence for it to be deemed personal data.[532] Moreover, it makes the point that generic references to social media and non-specific assertions that there is 'so much information out there' – i.e. a basic level of conjecture - are not sufficient to change the balance of risk from remote; nor is the claim that there are those who want to contact the trial participants.[533] Nevertheless, the FTT (and the IC) discounts taking into account the confidentiality agreement to be signed by (independent academic) recipients of the modified data – who would also go through an approval process before receiving the data - for its part in mitigating the re-identification risk to a remote level.[534]

---

identification". For example, the IC commented that the hypothesis that identification is possible by combining the patient data with NHS data was implausible as (assuming that they have the relevant skill and motivation) this would involve an NHS employee breaching professional, legal and ethical obligations. Ibid, p.39: *"[the] hypothesis that identification is possible through combining that information with NHS data (involving an NHS employee both having breached their professional, legal and ethical obligations and also having the skill and inclination to so do) is implausible"*.

[531] The Code, p.25 - the ICO comments that it is often acceptable to carry out "*a general assessment*" where big datasets are concerned, including "*the chances that those who might be able to re-identify are likely to seek out or come across the relevant data*". Albeit In the IC's decision (FS50565190), it was stated that it was not clear to the IC what prior knowledge an individual might have which might allow them to identify one or more of the trial participants from the information requested, particularly given the large number of participants.

[532] FTT Judgement, p.39: "*[i]n short, we accept and adopt the Commissioner's wider submissions and reasoning as set out in his Skeleton Arguments and Written Closing on this issue. In all the circumstances and on the evidence before us we are satisfied that the risk of identification has been anonymised to the extent that the risk of identification is remote"*; and p.42: *"encryption makes the chance of identification even more remote in any event and strengthens our view that the speculative assertions of the occurrence of possible events actually taking place in a way that could lead to identification of individuals, by Professor Anderson are indeed remote".* Compare, the IC decision (F FS50565190) where it was stated that it was considered *not reasonably likely*, on the facts presented, that another person could identify the trial participants from the requested information in combination with other information that may be available to them.

[533] To that end, the FTT concludes that the chance of third parties (including a motivated person with specialist skills) discovering other information which, together with the data requested, could lead to individual identification on the facts of this case was remote.

[534] See, Knight & Stalla-Bourdillon. (2016, The First-Tier Tribunal and the anonymisation of clinical trial data: a reasoned expression of Englishness…. which would have to be abandoned with the GDPR?, [online]): "*[i]n fact, the opposite was argued by the IC: QMUL's provision of the data to independent scientists by way of research collaboration according to the limitational terms of signed, confidentiality agreements was held to amount to an acknowledgement that anonymisation was effective. The IC added that, otherwise, upon disclosure, QMUL would be in breach of both the participant consent agreement and the principles under the DPA. Finally, as per s. 40 of FOIA, compliance with data protection principles is crucial. If there is an argument that the release of effectively anonymised datasets is not governed by the DPA (because it is not personal data as defined in section 1), incentives to transform the datasets using anonymisation techniques for the purposes of disclosure to third party is undermined. Moreover, when the DPA is held to apply, a legal basis (ground of justification as set out in Schedule 2 of the DPA) is required in order to carry out any type of processing of personal data, unless the transformation is considered to be an act of further processing that is deemed compatible with the initial processing. Yet, on the*

5.3.2 Critical analysis of using a (re-)identificatory-approach to non-personal data assessed against data protection's twin objectives

This sub-section considers how the analysis in this section (5.2) affects the initial conclusions drawn at the end of Chapter 3 regarding the effectiveness of an identificatory-approach (under the DPD/DPA, before sub-section 5.2.3 considers the changes under the GDPR) in being able to protect individual interests, and maximise interpretational legal certainty.[535]

This analysis includes consideration of developments in resolving any 'grey areas' of uncertainty so far highlighted, along with any new advantages/disadvantages, taking a diametrically-opposed (*re-identificatory*-approach) viewpoint guided by the same objectives.[536] Yet, it also recognises that – in contrast with data that has ever been subject to anonymisation processes - additional information capable of linking the data with the original identity of a data subject de facto once existed and will often still be in the possession of the data controller, so the degree of re-identification risk will tend to be higher than under the scenarios considered in Chapter 3.

As a starting point, the analysis suggests a general rejection of the idea that it is never possible to transform personal data into non-personal data legally, because of the argument that advanced analytics and big data enabling cross-referencing with other information (known/unknown) make this impossible. Moreover, a hard-line approach (where statistical data would remain personal data as long as the raw data exists somewhere, consistent with an objectivist perspective on identifiability, and potentially also an absolutist position) is often rejected (e.g. in the UK - by the ICO and the judiciary).[537]

---

facts of this case, QMUL did not obtain consent from trial participants to release the data to members of the public. Thus, we reach the perverse conclusion that, by holding the data to be non-personal data, this would leave QMUL potentially free to do what it liked with the data, which would increase the re-identification risk level in practice. Whereas, if it was deemed personal data, the re-identification risk level would be mitigated by the application of data protection rules in practice (obligations upon data controllers to keep personal data secure). After all, releasing personal data to members of the public without imposing any contractual (security) obligations on the recipients of the data is surely incompatible with the initial processing of such data undertaken for research purposes? Thus, rather than disclosure of the data for research purposes being constrained by the DPA, as the FTT(IR) argues, it is in fact more important – RATHER - that the hypothetical disclosure of the data for non-research purposes would mean that QMUL would be acting in breach of the DPA if the data were deemed personal according to the facts of the case! The greatest deterrent against re-identification risk is the application of the DPA itself!"

[535] To note, there is also a strong societal interest in the processing of large datasets for research, while such activity might be inhibited when there is a large degree of legal uncertainty over the effective de-identification standard in law (bearing in mind that the application of data protection rules is binary, they either apply in full or do not depending on whether the jurisdictional threshold in met).

[536] That is, personal data becomes non-personal data – and its processing no longer subject to data protection rules - where a relationship of identification can *no* longer be deemed to exist between the data and that person (it ceases to identify or be identifiable of that individual).

[537] See, for example, the case-law summary in Appendix 3, from which it appears that the predominant UK view currently is that the personal status of data that has been anonymised in statistical form should be subjectively-determined according to a relativistic re-identifiability approach under the DPA. On that basis, in the UK a data controller should be able to aggregate personal data as statistics and disclose them in modified form as non-personal

Nonetheless, this analysis is only a 'snapshot' to show that much uncertainty remains, illustrated by the range of different interpretive approaches to the identificatory-approach in an anonymisation context within and between EU MSs. In the UK, for example, much remains unsettled regarding interpretation on the key points (re-identification 'how', 'from what perspective', and 'with what likelihood'?). For example, the *QMUL* case suggests that in assessing what means of re-identification are likely reasonably used, there are supplementary limitations upon the motivated intruder test, and there should be tangible evidence that re-identification risk is more than a remote possibility. By contrast, in the Code the ICO refers to the mitigation of re-identification risk until it is "extremely remote", not just "remote". Indeed, it appears that there is confusion between the application of the DPD Means Test, and the degree of risk required for data to be non-personal under the DPA, a distinction drawn by the ICO in the Code.[538]

At the pan-EU level, there is no CJEU case law directly relevant.[539] While WP216 develops its analysis on the three types of (re-)identification risks for consideration (addressing the question of '(re-)**identification how?**'), some of the WP's statements suggest that data that has been pseudonymised should always be deemed personal data as long as the raw data and its key-code exists in the controller's possession. They suggest an objectivist approach (to the question, '(re-)**identification from whose perspective'**) only taking into account the data controller perspective. This approach undermines incentives for organisations to create and use data with a lower re-identification risk, such as data for subjecting to pseudonymisation as a privacy-enhancing technique. In turn, this approach also undermines the attainment of data protection's first objective (where sharing is carried out without such protective measures applied) and/or its second objective (if less data are processed for sharing due to the perceived risks of data protection law applying, or at least uncertainty about whether it would apply).[540] By contrast, other WP comments suggest an

data because others cannot identify living individuals from such data (although the data controller may continue to process the raw data as personal data if it retains it in compliance with the DPA).

[538] The Code, p. 12: "*[n]]ote that the UK's DPA is framed in terms of identification or the likelihood of identification. The Data Protection Directive refers to 'likely reasonably'. In some cases the UK courts have used the 'likely reasonably' test. However, the practical problems that arise are much the same whether the test is of 'likelihood' of identification or 'reasonable likelihood' of it".*

[539] Although, to note, it can be contended that a dynamic IP address is comparable to a pseudonymised form of personal data where an ISP holds the 'key' to identify the data subject. To that extent, the findings by the CJEU in the *Breyer* case may be considered analogously-relevant. Notwithstanding, on the facts of the case, the IP address 'key' was held by the ISP and not the data controller at issue (the governmental web operator), in contrast to a pseudonymisation scenario where the data controller usually retains the 'key' (raw data and decryption code).

[540] As mentioned, per Knight & Stalla-Bourdillon (2016, p.299): *"where the transformed dataset is still considered to be personal data in the hands of the subsequent recipient of the dataset, complying with the entire gamut of data protection obligation is likely to have a chilling effect on data sharing given the complexity of distilling the new compliance obligations that will be required to be followed by organisations under the GDPR. This could have, in particular, a significant deterrent effect on the carrying out of beneficial longitudinal research studies that rely upon the reuse of personal data once it has undergone anonymisation methods, such as in the health or education sectors".*

"optimal solution" is possible whereby re-identification risk levels in each case depends on assessing the most likely and reasonable means the data controller or third parties may employ.[541]

Another notable trend is the gradual acceptance of the practice of objectifying a motivated intruder as a tool to help assess re-identifiability from transformed data. This suggests that if a hypothetical third party, with certain attributes, can be deemed capable of identifying the data subject from the data, it should be considered personal data for all those who fit that 'profile' (including knowns as well as unknowns). Yet, the types and extent of the attributes/capabilities delimiting this perspective are also in dispute at the UK level, as well as the EU level. Therefore, even if accidental matching possibilities resulting in re-identification are discounted under such a view (because they are not likely, or reasonably likely),[542] there remains a problem of legal certainty in terms of assessing whether – against which parameters - a hypothesised third party figure type (known or unknown) might re-identify the data subject.[543] Even if there were agreement on this issue, questions remain over the degree of likeliness of re-identification risk that is required for Functional Legal Anonymisation status to be achieved in any one particular set of circumstances ('(re-)**identification risk to what degree**?').

---

[541] WP216, p.12: "*a solution against these three risks [singling out, linkability, and inference] would be robust against re-identification performed by the most likely and reasonable means the data controller and any third party may employ. The Working Party emphasizes, in this connection, that techniques of de-identification and anonymisation are the subject of ongoing research and such research has shown consistently that no technique is devoid of shortcomings per se".* See also, p.24: *"[t]he optimal solution should be decided on a case-by-case basis. A solution (i.e. a complete anonymisation process) meeting the three criteria would be robust against identification performed by the most likely and reasonable means the data controller or any third party may employ. Whenever a proposal does not meet one of the criteria, a thorough evaluation of the identification risks should be performed."* Notwithstanding, the WP says that the following "good anonymisation practices" to reduce re-identification risk should take into account – not only technical elements – but also the following (pp.24-25):

- not relying on the "release and forget approach" given the "residual risk of identification" (see also at p.24: ***data controllers should: o 1. Identify new risks and re-evaluate the residual risk(s) regularly, o 2. Assess whether the controls for identified risks suffice and adjust accordingly; AND o 3. Monitor and control the risks…****As part of such residual risks, take into account the identification potential of the non-anonymised portion of a dataset (if any), especially when combined with the anonymised portion, plus of possible correlations between attributes (e.g. between geographical location and wealth level data)*" (emphasis added));
- *"[t]he purposes to be achieved by way of the anonymised dataset should be clearly set out as they play a key role in determining the identification risk";*
- *"all the relevant contextual elements – e.g., nature of the original data, control mechanisms in place (including security measures to restrict access to the datasets), sample size (quantitative features), availability of public information resources (to be relied upon by the recipients), envisaged release of data to third parties (limited, unlimited e.g. on the Internet, etc.)";* and,
- *"possible attackers by taking account of the appeal of the data for targeted attacks (again, sensitivity of the information and nature of the data will be key factors in this regard)".*

[542] Whereas, having a good understanding of previous data releases, in addition to other data sources that are available, to determine the likelihood of re-identification, may not be realistic in many cases.

[543] In this context, questions that remain unaddressed similar to those found at the end of Chapter 3 include whether the Means Test should incorporate consideration of context-dependent motives to attempt re-identification. Despite the obvious connotations of the term 'motivated' intruder, for example, the ICO says that - although one should not strictly consider other factors beyond means of identification - other relevant factors for assessing can include the potential for financial gain from re-identification (the Code, p.23).

Thus, as with Chapter 3's conclusions, legal uncertainty persists under EU/UK law as to what is required for making a (re-)identifiability connection to a person in respect of information already deemed relating to him/her. Specifically, much of this uncertainty applies where assessment of the data context suggests antecedent conditions are present, factually, for a non-remote possibility of future identification to occur without this being a *significantly probably* event (i.e. in the very middle of the (re-)-identification continuum of risk). Part of this uncertainty relates to a lack of clear and objective standard upon which to base a binary distinction (per p.198 above, e.g., it is not clear whether and why (extreme) remoteness is meant to equal non-significance, or what relationship the two values have) and how this should apply in practice. For example, practically (and similar to the example discussed in p.157 above), the expense needed to be outlaid by an intruder to de-encrypt raw data to reveal a data subject's identity directly may seem extremely remote if the controller took expensive encryption precautions. However, in the same circumstances the likelihood of identification may be deemed significantly probable by an intruder where publicly-available additional information could allow them to easily identify a data subject indirectly on balance, especially where the raw data is not stored securely or the redacted data is wholly or partially available. Similarly, it is unclear exactly what EU law requires should be done, in releasing data that retains indirect identifiers, to ensure that, on disclosure, the data would, in the hands of third parties, lose its personal data character.

Of course, this discussion focuses our attention back to where law-makers have decided to posit the line between personal and non-personal data, as well as how policy-makers and relevant others have decided to interpret that line in practice, together with making consistent sense of different interpretations.[544] Moreover, this is not solely an EU legal issue. For example, under US law, an identificatory-approach underpins the definitions of PII in federal and state privacy laws, and related data anonymisation policy. For example, the Federal Trade Commission (FTC) in a 2012 Staff Privacy Report relevant to all industries, refers to the language of re-identification risk in discussions about non-PII.[545] Notwithstanding, the FTC explicitly endorses a standard (in satisfaction of which its privacy framework does not apply) of "reasonable (non-)linkability" in this Report. It can be satisfied, says the FTC, to the extent that a company: *"(1) takes reasonable measures to ensure that the data is de-identified; (2) publicly commits not to try to reidentify the data; and (3) contractually*

[544] See, European Commission (2003, p.13): p. 13: "*[t]here appears to be division among Member States on whether or not to use a relative approach to the concept of personal data in the sense that data are considered personal only for someone who can link the data to an identified individual.* ***The laws in some Member States make clear that for instance encoded or pseudonymised data are 'personal' with regard to a person who has access to both the data and the 'key', but are not personal with regard to a person without access to the 'key'"*** (emphasis added). For example, countries such as Belgium, Italy, France and Sweden have taken the view that in principle all data which can be linked to an individual is regarded as personal data irrespective of whether data is processed by someone who cannot make the link.

[545] Federal Trade Commission (2012).

prohibits downstream recipients from trying to re-identify the data".[546] Hence, US federal law/regulation neatly side-steps the issue of how to delineate precisely how much 'linkability' risk is reasonable through the use of safe harbour presumptions of legality assuming certain steps are taken by data holders. This issue is returned to in Chapter 6.

5.3.3 The GDPR

In reminder, Recital 26 GDPR states:

> The principles of data protection should apply to any information concerning an identified or identifiable natural person…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…** (emphasis added to indicate key changes from Recital 26, DPD)

Recital 26 goes on to say, *"[t]he 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 to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable"* (emphasis added). This interpretive emphasis via the inclusion of the word "therefore" suggests that the GDPR adopts the Functional Legal Anonymisation standard, requiring data controllers to undertake risk analyses in respect of what means may "reasonably likely" (changed from "likely reasonably" under Recital 26 DPD) be used by them and by third parties to re-identify an individual from data. To this extent, it re-confirms a re-identification risk-based - and dynamic - modelled concept of non-personal data.[547] Furthermore, the new reference to "*all objective factors*", with examples, may be viewed as providing further support for the trend of

[546] Ibid, p.iv. With respect to the first prong of the test, the FTC states that this *"means that a company must achieve a reasonable level of justified confidence that the data cannot reasonably be used to infer information about, or otherwise be linked to, a particular consumer, computer, or other device"* (p.21). Whereas, if a company does take steps to re-identify such data, its conduct could be actionable under Section 5 of the US Federal Trade Commission Act of 1914, which prohibits unfair or deceptive acts or practices in or affecting commerce.

[547] See Knight & Stalla-Bourdillon (2016, pp.287-288): "*[t]his dynamic state is epitomized by the fact that anonymized data can become personal data again, depending upon the purpose of the further processing and future data linkages".* By contrast (p.287): "*a static approach…tends to assume that once the data is anonymized, not only can the initial data controller forget about it, but also the recipients of the dataset are free from any obligation or duty because the transformed dataset always lies outside the scope of data protection laws".*

objectifying hypothetical possibilities of identifiability under the Means Test (similar to the effects of importing the motivated intruder test), even if such assessments remain context-dependent.[548]

Nevertheless, there is still a lack of clarity about what the Means Test encompasses and, indeed, further confusion created under the GDPR regarding a new legal definition it introduces. Article 4(5) GDPR defines 'pseudonymisation' as:

> [T]he processing of personal data i**n such a way that the data can no longer be attributed to a specific data subject without the use of additional information,** as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person" (emphasis added).

Recital 26 GDPR also refers to pseudonymisation in commenting on the effect it has on the legal status of data relating to persons: "*[p]ersonal data which have 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*". Yet, these GDPR quotes appear confused in comparison:

- First, under Article 4(5), data that has undergone pseudonymisation requires the keeping separate of additional information to ensure non-attribution (of data to an individual by a third party), so the second (Recital 26) quote - where attribution is possible - could be read as incompatible with this definition and otiose (at least from a third party perspective).

- Second, GDPR references to 'pseudonymisation' (including the two quotes above) could be interpreted as all referring to a process by which personal data becomes no longer attributable to a natural person without the use of additional information. However, while the Recital 26 quote suggests that it is the added-value of additional information usage – enabling attribution - that tilts the balance as to whether (the processing of) such data are subsequently deemed personal data or not, it does not explain why this is exactly in terms of the definitional detail mentioned in the Article 4(5) quote. For example, could it be because – despite arrangements in place to keep relevant additional information separate – the security of this arrangement, and/or other measures implemented to ensure non-attribution, is inadequate?

[548] To this end, it also provides more legal certainty in interpretation of the (re-)identificatory-approach the context of assessing when personal data have become non-personal data (while noting that the test remains dynamic as it is context dependent so ultimately a degree of legal uncertainty is unavoidable, as previously discussed).

Furthermore, a new point of confusion is raised regarding whether the Recital 26 quote should be interpreted in the same way as WP216's statement that if data controllers retain raw personal data at source, its modified form remains personal data.[549] As Knight/Stalla-Bourdillon comment:

> One way to make sense of this sentence would be to say that, as long as the raw dataset has not been destroyed, a transformed dataset must only be considered pseudonymised and remain subject to EU data protection laws…which… is not fully consistent with a risk-based approach to anonymisation.[550]

Hence, a new prima facie advantage of the (re-)identificatory-approach under the GDPR flows from the new definition of pseudonymisation, because it provides formal (albeit implied) acknowledgement of levels of re-identification risk. However, interpretational confusion gives rise to new disadvantages under the GDPR to add to those previously discussed in Chapter 3. In particular:

- **(Re-)identification how? -** There is no mention of linkability risk in the GDPR.
- **(Re-)identification from whose perspective?** - Recital 26 GDPR, like Recital 26 DPD, suggests a relative perspective for assessing re-identifiability under the Means Test. Yet unresolved is the problem mentioned above regarding agreement over the limitational contours of hypothesised third party figure type (known or unknown) by which they could re-identify a data subject.
- **(Re-)identification risk to what degree? –** A misconception currently tars the new pseudonymisation concept: that is, applying pseudonymisation to personal data is determinative in and of itself as to such data's personal status legally; whereas in reality

---

[549] WP216, p.9: *"it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data*". The only exception, the WP says, is where "*the data controller would aggregate the data to a level where the individual events are no longer identifiable*".

[550] Knight & Stalla-Bourdillon (2016, p.301). In other words: if pseudonymisation is read under the GDPR as meaning the process by which formerly personal data becomes no longer attributable to a natural person without the use of additional information; and, Recital 26 implies that personal data can include data that has been subject to pseudonymisation as long as it "*could be attributed to a natural person by the use of additional information*"; therefore, logically, data that has been subject to pseudonymisation should be deemed personal data as long as the raw dataset remains in existence through which attribution can be made (even if the route of anonymisation through aggregation has been chosen). The crux of this confusion is rooted in the fact that the GDPR does not refer to linkability in its definition of pseudonymisation at Article 4(5). The alternative interpretation is that there does remain a route to effective anonymisation through aggregation, because emphasis should be placed on the identificatory capacity from combining information only, which Knight/Stalla-Bourdillon believe to be the "*most sensible path*" (ibid, p.301): *"[t]his interpretation makes better legal sense as the removal of individual-level elements within a shared dataset truncates in principle the possibility of any harm befalling to individuals through the linking of individualised data records from which they could be singled out*". This is because the GDPR at various places encourages the use of pseudonymisation and/or notes the benefits for organisations in using pseudonymisation as a technique to lower privacy risks of processing personal data to the data subjects concerned (Recitals 28-9, 78, Articles 6, 25, 89).

data that has undergone pseudonymisation will only be personal data assuming that other measures have not been taken to mitigate re-identification risk to a safe level.[551] Said otherwise, those intending to process data that has been subject to pseudonymisation should assess these and other measures being used, in order to establish whether the data would be subject to data protection principles or not. Under the GDPR, this message appears clouded.[552]

5.4 The need for effects-based exemptions to complement the identificatory-approach to the concept of non-personal data under the GDPR

This section combines evidence that there is room for an effects-based exemption model in the GDPR data protection regime that remains identity-based, together with a taster of certain advantages its introduction would have in the evolved case study discussion (before Chapter 6 addresses the issue of how in more detail).

Assessment remains against the standard of how best it might achieve data protection's twin objectives in that context (while maintaining broad coherency with existing data protection terminology/principles). There is also consideration of incentives and development of the argument that an effects-based exemption model might provide additional benefits in encouraging data controllers to engage with jurisdictional issues in more than a perfunctory one-off manner (as is

---

[551] Compare the static implications of the European Parliament's proposals (in its legislative resolution of 12 March 2014, P7_TA(2014)0212) to add a definition for 'pseudonymous data' as a new Article 4(2)(a) into the GDPR: *"personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution*" (amendment 98 to Article 4 (insertion 2(a))). To note, the European Parliament's Committee for Civil Liberties, Justice and Home Affairs ('LIBE') – which leads the European Parliament's legislative deliberations on the proposed data protection reforms - produced a report on the Draft Regulation endorsing this approach (European Parliament Committee on Civil Liberties, Justice and Home Affairs. Draft report on the draft Regulation PR\922387EN.doc, 17 December 2012). It also suggested alleviations with regard to legal obligations placed upon data controllers in respect of such definition-satisfying pseudonymous data compared to anonymous data (p.214). For example, it proposed that where personal data is processed only in the form of pseudonyms, the data controller may be able to obtain the data subject's consent to that processing through automated means using an EU technical standard (p.76).

[552] See, for example, Cheshire. (2017, Tfl plans to make £322m by collecting data from passengers' mobiles via Tube Wi-Fi, [online]) and the two quotes from privacy experts at the bottom of the article fixating on the implications of the term 'pseudonymisation': "*TfL worked with the Information Commissioner's Office on the scheme and said that user data was anonymised. But privacy experts have cast doubt on the implementation. Paul-Olivier Dehaye, the cofounder of PersonalData.IO, told Sky News: "TfL don't seem to understand what 'anonymised' means in data protection terms. While the pilot was running, the data was merely pseudonymisation, while retaining the technical capacity of easily combining this data with external datasets. "In essence, the value and dangers of this data are still fully there, but TfL has merely constructed a fiction that the individuals were not identifiable and conveniently assumed that would free them from the legal safeguards." Dr Lukasz Olejnik, independent cybersecurity and privacy researcher, told Sky News: "TfL has definitely identified some privacy risks and tried to tackle them. They should be applauded for that. "It's important to note that TfL does not provide an anonymization scheme. It's called pseudonymization, as the data are not processed in a way making it impossible to calculate the data back, given resources*".

often the case now when it comes to considering data from which indirect identifiers have been removed).

This analysis surpasses the Effects-based Approach narrowly conceived in Chapter 4, which has been rejected. For completeness, however, in Appendix 4 the reader can find a discussion of the advantages and disadvantages of using the Effects-based Approach for determining when to deem data sufficiently non-personal to fall outside of data protection rules (in particular, by considering an assessment framework that could be employed by data controllers when deciding whether personal data exists in particular circumstances).

5.4.1 Associations between anonymisation and privacy harm mitigation

Per Chapter 4, risk-based discourse is embedded in many ways in current law and the GDPR, but to what extent can this be said to be a risk-of-harm (negative effects) discourse implicitly, rather than a re-identification risk issue, when it comes to the concept of non-personal data?

The answer is to a great extent. As mentioned, anonymisation and privacy have been closely linked by the courts, policy makers and in academic literature. Authoritative regulator support for an effects-centric discourse under EU/UK data protection law in the context of determining the status of data that has been subject to anonymisation techniques is set out in Appendix 4 below.

In legal scholarship, associations are also aplenty. For example, Floridi points out that there is a perception that "*privacy concerns raised by Big Data practices can be addressed merely by removing identifying information*".[553] Policy-makers and others often share this perception, linked to the explanation of an associated dilution in the potential for privacy impact post-anonymisation. Polentsky/Wolf, for instance, intimate that personal data de-identified effectively do not raise significant privacy concerns because the likelihood of individuals being harmed from its use has been mitigated:

> **Anonymizing personal information decreases the risks that personally identifiable information will be used for unauthorized, malicious, or otherwise harmful purposes. Properly anonymized data are highly unlikely to have any impact on individuals and do not implicate privacy concerns.** When data sets are anonymized properly, re-identification is no easy task.[554] (emphasis added)

[553] Floridi & Mittelstadt (2016).

[554] Polentsky & Wolf (2013, pp.7-8). To note, while this quote associates the risk of privacy harm flowing from the use of data with re-identification risk, as previously mentioned, the two are not necessarily synonymous. This is most clearly the case when a narrow concept of identification (based upon 'to identify' interpreted narrowly as a verb linked to 'real-word' identity as discussed in Chapter 1) is used.

Other scholars recognise a distinction between the application of anonymisation processes in reducing re-identification risk, and the reduction of the risk of harm (negative effects), notably as follows (emphases added):

- Oswald describes anonymisation as a two-part assessment of "*the possibility of something bad happening; i.e. the likelihood of someone being able to re-identify an individual, **and the** harm or impact if that re-identification occurred*".[555]

- Mantelero states, regarding anonymised information, "*given the impossibility to exclude any risk of re-identification, the adopted solutions should be proportional to the risks of re-identification that exist in specific data processing **and to** the related potential negative impact on data subjects*".[556]

- As mentioned above, Hon et al also argue that – while the definition of personal data "*should be based on a realistic risk of identification*"[557] – the "*applicability of data protection rules should be based on risk of harm and its likely severity*".[558]

5.4.2 Arguments for doing more to encourage the benefits the re-usage of data relating to people (by third parties) under data protection law

These and other scholars also recognise that, typically, anonymisation techniques are applied to personal data to allow its subsequent use post-modification and, specifically, to help maximise data value while minimising the risk of negative impact upon individuals when such subsequent processing occurs. Furthermore, optimally, the legal framework would promote this outcome and

[555] Oswald (2013, p.24).
[556] Mantelero (2015), who also says: "*[i]n this sense, the nature of the database (quantity and quality of information collected), the risks of illegitimate access to information (physical and logical protection of databases, internal procedures adopted by data controller), and the resources that are necessary for reidentification (time, equipment, etc.) represent the factors that should be taken into consideration to identify the most adequate level of anonymization".*
[557] Hon et al (2011, p.224). Compare p.226: "*[w]here identification risk is remote or highly theoretical, for example due to technical measures taken, we suggest information should not be 'personal data'. In particular, encrypted data should be recognized as non personal data in cloud computing, at least where strongly encrypted. Clarification is also needed regarding anonymized data and anonymization and encryption procedures. The current law partly recognizes this ('means likely reasonably to be used', and WP136's reference to theoretical risks). However, in today's environment it may make sense for the threshold to be higher, based on realistic risks (such as 'more likely than not'). The boundary should be clearer*".
[558] Ibid, p.211. Compare p.224: "*[w]e suggest there should be a two-stage approach. First, the definition of 'personal data' should be based on the realistic likelihood of identification. Secondly, rather than applying all the Principles to information which has been determined to be 'personal data', there should be consideration in each particular context of which data protection rules should be applied, and to what extent, based on the realistic risk of harm and its likely severity";* and p.225: *"rather than considering solely whether information is personal data, it may make more sense to consider risk of identification and risk of harm to individuals from the particular processing, and the likely severity of any harm. Processing should then be tailored accordingly, taking measures appropriate to those risks."*

facilitate data sharing, and so encourage the realisation of potential benefits from the secondary processing of information about people to which access has been provided.[559]

Managed effectively, for example, the unexpected insights from advanced analytics on large datasets (accumulated from multiple sources – such as linked across sectors and applications - collected for distinct purposes) might offer substantial benefits, such as improved public service delivery. Anonmyisation techniques continue to play a significant role in this respect.

However, arguably the GDPR does not provide sufficient incentives taking into account the impact of the development of big data. Something else is needed to fuel a shift towards more sharing of data relating to people, in particular to promote analytic techniques on data in innovative re-usage ways, to help deliver the benefits of big data (economically, societally, and to individuals as consumers and citizens), ***and*** ensure privacy rights are protected adequately. In other words, setting up a contest between choosing the positive effects of big data at the expense of creating/increasing negative effects (to an appreciable level) is not necessary.

The challenge, rather, is to be creative in coming up with new and practical ways to create incentives to those holding information to share it with those who would repurpose data about people for innovative ends to more readily realise potential benefits of big data, and adequate data protective measures to mitigate the risk of appreciable harm, side by side. How can we get closer to a regulatory outcome that enables the uncovering of value in data to create new economic/social value? In other words, we need data protection mitigatory (privacy preserving) measures seen as utility beneficial – which can be proactively not just defensively to facilitate data innovation. This would mark a move towards a more pro-active data protection policy in favour of socially beneficial outcomes.

5.4.3 Evolved case study to introduce the benefits of effects-based exemptions

Let us revisit again the case study discussed in Chapters 1 and 4 and introduce an altered set of facts, which would allow different insight into the virtues of the effects-based exemption proposition now being introduced for proposal as a regulatory compliance-facilitating tool. In this

[559] See, e.g., Tene & Wolf (2013, p.4): *"[o]ptimally, the legal framework would allow for use of data in ways that maximize such value while minimizing privacy risk. Yet over the past decade, it has become clear that in a world of big data, de-identification based solely on technical measures is increasingly strained by the existence of increasingly effective re-identification techniques. Today, ample evidence exists of re-identification of apparently de-identified data, through methods such as multiple queries on a single database or linking the data in one database with that in another, which is often publicly available online. Moreover, measures commonly taken to de-identify personal data, such as scrubbing or aggregating certain data fields, degrade the utility of the data for future use. The result is that strictly technical de--identification has become encumbered with a deadweight cost, reducing data utility without eliminating privacy risks".*

context, it is worth re-emphasising that there is strong societal interest in the processing of large datasets for pure research activities, whereas such activities might be inhibited when there is a large degree of legal uncertainty over the effective de-identification standard in law (and, not forgetting, the application of data protection rules is currently binary, they either apply in full or do not depending on whether the jurisdictional threshold in met).

Let us assume that the Trust wanted to pseudonymise and then share the patient-related health data (current and historic) to DeepMind as a data controller (joint or otherwise with the Trust) in its own right to carry out advanced analytic processes on the data - of the machine-learning/AI technology type that DeepMind is famous for - for pure research purposes. In other words, let us assume that DeepMind would have considerable discretion to determine the purposes and manner in which real personal data from a live data subject, which had been pseudonymised, would be transmitted to it for secondary processing relating to a specific research study in a closed research environment. This research would focus on the use of machine-learning techniques to develop better algorithms – to detect and learn patterns from data quickly - that could lead to earlier detection of AKI than traditional diagnostic tools.[560] Compared to traditional methods reliant upon human analysis - unable to compare a new patient's information with millions of other cases – such projects (including providing technology platforms that enables analytics as a service for NHS Hospital Trusts) might be able to predict when individuals are in the early stages of a disease that has not yet become symptomatic. Thus, the motivation behind the data sharing is to enable DeepMind's software to be developed able to identify patterns that can assist doctors and nurses in treating patients or in recognising conditions earlier than normal, resulting in running tests upon a particular individual when a pattern alert is triggered to check if the prediction is correct.

Let us also assume that DeepMind agrees with the Trust to use the legal mechanism of a data sharing agreement – setting forth the permissible conditions under which DeepMind may access

[560] To note, the Trust and DeepMind did in fact enter into a research-related Memorandum of Understanding on 28 January 2016. The Memorandum of Understanding primarily relates to a proposed AKI project involving research for project development on pseudonymised data under the National Research Ethics framework. However, that research project was subsequently abandoned. DeepMind has, however, made public other partnerships with the NHS, for research rather than patient care, with actual involvement of AI, and with appropriate research approval. One, with Moorfields Eye Hospital in London, involves DeepMind receiving one million eye scans which it will run through its machine learning algorithms in search of new patterns of macular degeneration that might allow disease to be caught earlier. Approval to work on pseudonymised data was given on June 2016 and the project was announced publicly the next month. For more, see project aimed at using machine-learning techniques to improve diagnosis of diabetic retinopathy (Hearn, A. Google DeepMind pairs with NHS to use machine learning to fight blindness. [online] The Guardian, 5 July 2016, available at: https://www.theguardian.com/technology/2016/jul/05/google-deepmind-nhs-machine-learning-blindness [Accessed 1 August 2017]). DeepMind has also engaged in other big data projects using algorithms to help public health initiatives, such as aimed at helping radiotherapists treat head and neck cancers. For more on the 'Routes to diagnosis' big data project to use algorithms to analyse over 100 million patient records, see Hearn, A. Google DeepMind and UCLH collaborate on AI-based radiotherapy treatment. [online] The Guardian, 30 August 2016, available at: https://www.theguardian.com/technology/2016/aug/30/google-deepmind-ucl-ai-radiotherapy-treatment- [Accessed 1 August 2017].

and use the shared data, as well as restrictions on further sharing with other parties. Moreover, DeepMind agrees it will implement appropriate technical and organisational measures to ensure that it stays in control of the level of agreed access and (re)usage based on the specific circumstances of the planned processing activities.

It is in this hypothetical scenario that the benefits of introducing a new effects-based exemption model in practice can be illustrated. It is possible to conceive of the introduction of a research-specific block exemption regulation (hereafter 'a Research BER') to apply automatically to data processing for pure research purposes, which could help promote this beneficial activity in a legal framework giving more trust to all parties and with advantages over (existing and GDPR) status quo.

A research BER is chosen because, as alluded to above and in Chapter 1, of the potential benefits that big data analytics research is helping to deliver in ways that improve the quality of people's lives. The availability and use of large data sets for research purposes directly benefits society. It enables scientific breakthroughs, commercial innovation, and improvements in government services including health, education, transportation, housing, and public safety. Health data has huge potential for medical research and there is public interest in the secondary usage of data by the private sector.

Those intending to process data relating to people that are able to satisfy the conditions under such a Research BER would then have the opportunity to obtain the benefit of an automatic formal legal presumption that they are exempt from data protection rules (in effect, because they fall within a legitimising 'safe harbour' providing certainty of non-infringement of data protection rules in respect of the processing to which it applies). Critical to the application of this model is the importance of harm-mitigating conditions that organisations can would aim to satisfy (and demonstrate as such) to fit within the ambit of this new block exemption. Such conditions would be based upon a likely-effects based analysis related to processing activities carried out in certain contexts.

Legal precedence for this block exemption model embedded into a legal instrument endorsed by regulators, as well as the multi-factorial effects-centric methodology that would sit behind setting the criteria contained within them to guide organisations linked to the overall context of processing, are set out in the next chapter. This condition-setting would be based on the regulator's well-grounded understanding and interpretational guidance regarding the mechanics of the processing harm to be avoided. For now, suffice to say the availability of the automatic safe harbour on offer under the Research BER would encourage organisations to take steps (and continue to ensure they are upheld) aimed at reducing the potential for harm flowing from their planned processing activities in their research context.

The need for such an instrument becomes greater when considering data sharing with private companies, where there is tension (or is at least perceived to be tension between) potential benefits to such companies using big data and the privacy of a data subject about whom data is being shared for secondary reuse. While people are generally supportive of the use of data for research provided there is a clear public benefit, trust-related concerns are often expressed in that context where private companies are involved about inappropriate usage.[561] The question arising is: how to minimise data risks while maximising benefits to all parties?[562]

One answer to help fill that trust 'gap' is to boost affected people's confidence that the law is being followed and provides sufficient protection, twinned with incentives for compliance. Thus, it could be a condition of a Research BER that data controllers demonstrate and make transparent the benefits of secondary data usage (e.g. by providing easy-to-understand summaries of the potential contributions of research outputs from studies to which individuals' data have contributed). Furthermore, the types of benefits possible could be a key focus of a Research BER, by explicitly discussing the possibility for collective benefits as well as describing the particular processing conditions under which different types of benefits are most likely to arise.

Of course, the DPD/GDPR do explicitly recognise the benefits associated with personal data processing in a research context, including the potential for new knowledge about "widespread medical conditions" and the "long-term correlation of a number of social conditions." Recital 157 GDPR, for example, states that research results can "*provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people, and improve the efficiency of social services*". It is also acknowledged that the DPD (see discussion in the next chapter, at fn.644) and the GDPR explicitly provide some research exemptions - i.e. it might actually indirectly facilitate towards the outcome of the model being proposed. Specifically, Article 89 GDPR accommodates some flexibility – albeit at a MS's discretion – in relation to the non-application of certain data protection obligations in a research context where specific safeguards

[561] For example, concerns about data misuse are more likely to be expressed with respect to access by pharmaceutical companies, with patients less willing to share information in this context.

[562] See, e.g. House of Commons Science and Technology Committee (2016, p.5): *"[t]here are arguments on both sides of this issue: Seeking to balance the potential benefits of processing data (some collected many years before and no longer with a clear consent trail) and people's justified privacy concerns will not be straightforward… to strike a transparent and appropriate balance between those benefits and privacy concerns… in general, modifying the data to protect against attribute disclosure means reducing the plausible inferences that can be drawn from the data. This can be detrimental to the objective of learning as much as possible from the data and building generalizable statistical models from the data. Furthermore, to protect against attribute disclosure, one must anticipate all inferences and make data modifications to impede them, which may not be possible. Some inferences may be desirable because they may enhance understanding of the treatment benefits or safety of a new drug or device, and some inferences will be stigmatizing to the data subjects. Legislation should be kept to the minimum required to facilitate the uptake of the Internet of Things. It should enable more efficient public and private services in areas such as healthcare, energy and transport, and should aim to minimise threats and harms".*

are met (e.g. in respect of certain data subject rights, and an assumption of purpose compatibility when further processing is carried out for research purposes).

Nevertheless, the effectiveness of the GDPR in this area remains incomplete in terms of addressing the concerns raised in the revised scenario above: how to encourage data sharing to achieve benefits from research using data analytics. For example, while the GDPR adopts a "broad" definition of research, encompassing the activities of public and private entities alike (Recital 159), in an age where the big data analytics activities of many organisations may qualify as research, it is unclear exactly how far the GDPR's scientific research exemptions will extend. This is important as some exemptions (e.g. an exemption to the right to object to processing under Article 21 GDPR for research purposes) are only available where a processing task is justified by public interest (Article 21(6)) with Recital 45 GDPR specific that this means that it "should have a basis in Union or Member State law". By contrast, those doing research without a public interest (i.e. mandate) laid down in law would be ineligible. In other words, the GDPR research exemptions do not necessarily help those general analytics research activities increasingly conducted beyond the reach of traditional academic oversight. Furthermore, as described in the case study when revisited in Chapter 6, the legal analysis is complicated by the fact that if the research exemption does not apply, and when consent cannot be obtained for secondary data usage, the recipient has the extra challenge of demonstrating a new lawful basis for its processing that may not easily be achievable.

For now, arguably more can be done to map out the existing exemption model in a more thought-out way from an effects-based perspective, in particular to enhance collaboration opportunities between researchers and privacy sector entities seeking to use general analytics on real data to the public good. A Research BER could guarantee legitimacy for both data controller sharing activities, but also – as will be argued in the next chapter – the recipient's data processing activities could also be presumed legitimate on this basis. In particular, if the data for sharing has been pseudonymised - as in this case scenario, which in turn raises the question whether GDPR applies in the first place (hence the issue being fundamentally jurisdiction) - there is arguably more justification for introducing more scalable/flexible data protection obligations in light of the benefits that could be attained.

This discussion continues in the next chapter elucidating on the block exemption effects-based model and how it might reduce the compliance burden on controllers thinking about sharing data to a third party, and the third party receiving such data, for secondary beneficial processing purposes. At the same time, adequate safeguards sufficient to mitigate the risk of harm resulting from the re-usage of data relating to individuals would still need to be assured, in particular to

protect those individuals whose data are analysed to develop algorithms and, consequently, about whom future decisions could be made to their detriment.

5.5 Chapter conclusion

This chapter drew interim conclusions to the ultimate research question, drawing together the analyses of this and the last two chapters in respect of sub-research questions 1-2. These conclusions were restated as relevant to justifying the need for sub-research question 3 – aimed at developing a more robust effects-centric block exemption model, which may be considered consistent with the underlying philosophy of EU data protection law and effectively further both its goals.

Chapter 6 tackles sub-research question 3, developing the effects-based exemption model practically, aligned with the existing regulatory system and the GDPR, to boost the achievement of the twin aims and provide more certainty to those involved in data sharing scenarios where the potential effects of recipients' processing activities could also be taken into consideration. Rather than introducing wholly new concepts, the hope is rather to leverage and integrate incoming GDPR principles/obligations (including accompanying policies and mechanisms for maximum effectiveness). Moreover, in focusing on a new jurisdictional approach related to the possibility for an automatic presumption of legitimacy using at core an effects-based analysis, it will be argued that it can provide an important framing functionality that can be dovetailed to both substantive data protection principles and the new requirement to carry out DPIAs in certain cases under the GDPR.

Chapter 6 - Developing an effects-based exemption model inspired by an existing effects-based EU law and regulatory regime

This chapter develops potential effects-based approach models in response to the disadvantages described at the end of the last chapter and for those reasons indicated. Specifically, this development relates to the proposal that, when persons intend to process data relating to a particular individual and prior to that event, they shall carry-out an assessment of the likely harm flowing from the envisaged processing activity upon that individual to determine if data protection law will apply.

This modelling development draws inspiration from another legal/regulatory regime that focuses on effects-based assessments – EU competition law and policy – in considering ways to improve the effectiveness of an effects-based approach. This aim was formulated bearing in mind, "*the development of any robust theoretical framework capable of justifying and supporting decisions about the boundaries and definition of personal data' cannot take place within a vacuum*" but "r*emain grounded within the realities of practice*".[563] Specifically, analysis of model strengths and their potential improvement (as well as weaknesses) must be considered against the backdrop of how the EU data protection regulatory regime works in practice now and from 25 May 2018.

To this end, the practical disadvantages described in relation to the Effects-based Approach may perhaps be remedied through adoption of certain aids-to-compliance (tools) of the type used in competition law regulation, specifically those emphasising organisational self-assessment and reliance on legal safe harbour instruments. The latter are provisions that provide legal immunity of one type or another to those who fall within their ambit, typically involving organisations that want to avail themselves of this protection self-certifying periodically that they comply with a set of conditions within a relevant legal instrument in which the safe harbour is implemented. Therefore, this examination is carried out according to the terms of sub-research question 3:

> **Can the practical disadvantages associated with an effects-based approach to personal data be ameliorated by the use of block exemption provisions, as exemplified under EU**

[563] Booth et al (2004, p.100).

competition law (a distinct area of law and regulatory system that has been modernised using an effects-based policy approach)?

Responding to this question helps draw an overall conclusion in Chapter 7 on whether introduction of an effects-based jurisdictional model of some sort – in particular one introducing a framework of effects-centric exemptions - would be a marked improvement over the current identificatory-approach to the personal data concept and, if so, why. It also offers some illustration of novel thinking about how a new approach could address the shortcomings of existing identificatory models.

In terms of chapter structure:

- **Section 6.1** introduces the basics of EU competition law and describes how its 2004 modernisation resulted in certain regulatory advantages facilitated by emphases placed on effects-based, self-assessment requirements framed by certain legal safe harbour instruments.

- **Section 6.2** proposes and compares two possible evolved, alternative effects-based models (**Model 1** and **Model 2**) aimed at remedying the two practical disadvantages highlighted at the end of the previous chapter. These relate to strengthening regulatory incentives to assess the applicability of data protection law - and, especially, assessment of the likely negative effects (harm) likely flowing from an intended processing activity upon the person(s) to whom the data relates, and mitigation of such effects where they are deemed appreciable - as well as improving legal certainty.

- **Section 6.3** develops the preceding case scenario applied to the preferred evolved, effects-based exemption model (**Model 2**) if adopted in law, to illustrate its benefits.

- **Section 6.4** considers outstanding limitations regarding Model 2 that still need to be addressed, as well as possible future research questions that arise from its proposition concerning potential for a more risk-proportionate data protection regulatory system.

- **Section 6.4** concludes this chapter in highlighting the outline of a final response to the overall research question of this thesis, which is addressed in the final chapter.

6.1 Insights from an effects-based regulatory regime: EU competition law

Why competition law? For background, parallels have already been drawn between data protection law and competition (also known as antitrust) laws,[564] including by the influential EU Data Protection Supervisor (EDPS) who (as mentioned in Chapter 2) regulates EU institutions' data protection compliance. In speeches, the EDPS has said that data protection – a relatively new discipline - has "*a lot to learn from the more established legal regime of enforcing competition rules*".[565] Furthermore, the EDPS points out that they share twin objectives: "*EU principles and rules on data protection, competition and consumer protection have been designed to promote a thriving internal market and to protect the individual"*.[566]

Areas that the EDPS highlights as ones that data protection can learn from competition law include regarding sanctions (now strengthened under the GDPR), but also the principle of accountability – which, as previously discussed, essentially refers to the obligations that organisations have to follow in order to demonstrate compliance - as a cornerstone of the EU competition law regulatory regime and fundamental focus under the GDPR. To this end, the EDPS commented in 2014 regarding data protection law, "*the focal point for privacy should shift from users to (a) policymakers or self-regulatory leaders to determine the contours of accepted practices; and (b) businesses to handle information fairly and responsibly"*.[567] In 2015, moreover, the EDPS recommended that data protection law *"take a leaf from the EU's competition manual, where a relatively limited body of secondary legislation…encourages a culture of accountability and awareness among undertakings".*[568] The WP has also made comparisons between the competition law and data protection law, suggesting that the latter is capable of protecting individuals from having their rights and interests unduly prejudiced by organisations in the same way that the former does.[569]

---

[564] See, e.g., Tene (2015, Privacy Is the New Antitrust: Launching the FTC Casebook. [online]), and Costa-Cabral & Lynskey (2017). The latter paper (p.12) refers to "*calls for a broader structural approach to personal data protection by fostering closer integration between data protection and other EU law policies that shape markets, including competition law*", which the authors describe as indicative of the underlying significant 'family ties' between the two areas of law.

[565] See, e.g. a speech by Giovanni Buttarelli (Antitrust, Privacy and Big Data, 3 February 2015, [online], p.3 and p.11: *"[s]eparate rules on data protection, competition and consumer protection all converge around a two-fold purpose – the protection and promotion of the welfare of the individual and the facilitation of the creation of a single European market"*).

[566] See also e.g. EDPS (2014, p.6). This opinion, regarding privacy and competitiveness in the age of big data, examines the legal background to data protection, competition law and consumer protection and explores the interfaces between these areas in the digital economy.

[567] Ibid, fn.30.

[568] See EDPS (2015, p.3).

[569] See WP217 (p.40): "*[t]he status of the data subject and the data controller is also relevant when assessing the impact of the processing. Depending on whether the data controller is an individual or a small organisation, a large multi-national company, or a public sector body, and on the specific circumstances, its position may be more or less dominant in respect of the data subject. A large multinational company may, for instance, have more resources and negotiating*

Such comparisons specifically include allusions to the fact that the competition law regime places emphasis on organisational self-assessment (of its own intended conduct to ensure legal compliance),[570] including regarding the likely effects of certain commercial activities. An obvious comparison unfolds with the proposal to refine the effects-based approach into an effects-centric jurisdictional model using exemptions that works effectively in practice under data protection law. This, too, could be framed in such a way as to help shift the onus of ensuring compliance with its rules more firmly upon organisations (rather than relying predominantly on DPAs to police them).

6.1.1 Introduction to EU competition law

Competition law concerns economic activity and the exercise of market power. It has long been of central importance to the EU and its treaties, as its central aims are to enhance the efficiency of the Single Market, but also to ensure the welfare of, and choice available to, consumers.[571]

The legislative framework of EU competition law rules under the TFEU is broken down into three main pillars, with the last area being the one focused upon in this thesis:

- **Article 102 TFEU**, which prohibits the abuse by one or more undertakings of a dominant market position within the EU (or a substantial part of it) in a way which may affect trade between EU MSs;

- **EU Merger Regulation***,* which regulates certain mergers and joint ventures by companies operating in the EU according to a standard of whether they will "significantly impede effective competition"; and,

- **Article 101(1) of the TFEU**, which prohibits anti-competitive *agreements* between 'undertakings' (i.e. economically-active organisations), *decisions* by associations of undertakings, or *concerted practices* - which may affect trade between EU MSs - and which have as their object *or effect* the prevention, restriction or distortion of competition within

power than the individual data subject, and therefore, may be in a better position to impose on the data subject what it believes is in its 'legitimate interest'. This may be even more so if the company has a dominant position on the market. If left unchecked, this may happen to the detriment of the individual data subjects. ***Just as consumer protection and competition laws help ensure that this power will not be misused, data protection law could also play an important role in ensuring that the rights and interests of the data subjects will not be unduly prejudiced****"* (emphasis added).

[570] Ibid, fn.20.

[571] The CJEU, for example, has defined the concerns of competition law to be consumer welfare, the interests of competitors, and the structure of the market (Joined Cases C-501/06P, C-515/06P and C-519/06P, GlaxoSmithKline Services and Others v. Commission, [2009] ECR I-9291, para.63).

the EU.[572] Such agreements are legally unenforceable except where an exemption applies.[573]

In each of these three areas, an emphasis on assessing and evidencing likely competitive effects has become increasingly important in recent years. This trend was spearheaded by the European Commission post-2000 as it sought to develop a more rationalisable competition policy based upon economic analysis, alongside a more effective regulatory regime (so-called competition law 'modernisation', explained next).[574]

In brief, for Article 101(1) TFEU to apply three requirements must be satisfied, and the last two are subject to further delimitation as the effects must be deemed likely 'appreciable' according to the European Commission:

- There must be some form of agreement, decision, or concerted practice between undertakings;
- Which may *affect* trade (*appreciably)* between EU MSs; and
- Which has as its object or (*appreciable*) *effect* the restriction, prevention, or distortion of competition within the EU.

Notwithstanding, Article 101(3) TFEU provides for the possibility of exemptions from the Article 101(1) prohibition if certain criteria are satisfied. Briefly, the negative effects to which an agreement, decision, or concerted practice (hereafter, for ease, 'agreement') is likely to give rise must be evidenced as likely outweighed by corresponding beneficial effects.[575]

Therefore, this area of competition law places key importance on the assessment and evidencing of effects likely to result from certain agreements as pivotal in determining whether such activities

---

[572] TFEU, Article 101(1): *"[t]he following shall be prohibited as incompatible with the internal market: all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the internal market, and in particular those which: a) directly or indirectly fix purchase or selling prices or any other trading conditions; (b) limit or control production, markets, technical development, or investment; (c) share markets or sources of supply; (d) apply dissimilar conditions to equivalent transactions with other trading parties, thereby placing them at a competitive disadvantage; (e) make the conclusion of contracts subject to acceptance by the other parties of supplementary obligations which, by their nature or according to commercial usage, have no connection with the subject of such contracts".*

[573] TFEU, Article 101(2): *"Any agreements or decisions prohibited pursuant to this Article shall be automatically void".*

[574] Knight (2011, p.80).

[575] TFEU, Article 101(3): "*[t]he provisions of paragraph 1 may, however, be declared inapplicable in the case of: - any agreement or category of agreements between undertakings,- any decision or category of decisions by associations of undertakings, - any concerted practice or category of concerted practices, which contributes to improving the production or distribution of goods or to promoting technical or economic progress, while allowing consumers a fair share of the resulting benefit, and which does not: (a) impose on the undertakings concerned restrictions which are not indispensable to the attainment of these objectives; (b) afford such undertakings the possibility of eliminating competition in respect of a substantial part of the products in question".* Object restrictions, by comparison, are not eligible for such exemption.

are legal, or not (specifically, because the effects are not deemed appreciable). While this assessment can be carried out ex-post facto by a national regulator – competition authority - or courts (once harm has actually occurred), it is mostly emphasised as ex-ante (preventative) for carrying-out by companies before they enter into agreements based on the likelihood (and associated quantification) of negative effects that might occur in the future flowing from the intended business arrangement. It is also promoted as an on-going duty of self-assessment, requiring legal compliance (i.e. changing-environment-driven, risk-assessing) checks to be carried out periodically by relevant organisations.

6.1.2 Effects-centric modernisation of EU competition policy and enforcement

In 2004, EU competition law and policy was reformed, resulting in substantial changes to the procedures governing enforcement of its rules and regulator roles in this respect. Specifically, modernisation saw the entry into force on 1 May 2004 of Regulation 1/2003 on the implementation of the rules on competition laid down in Articles 101 and 102 of the TFEU (previously Articles 81 and 82 of the EC Treaty).[576] It introduced a 'directly applicable' legal exception system, meaning that organisations need to ensure themselves that their agreements do not have an appreciable negative effect on competition under Article 101(1). In case they think they do, organisations must try to obtain an exemption by meeting the Article 101(3) legal exemption criteria and, therefore, escape the negation of their agreement and its consequences, such as penalties imposed by national competition authorities on the companies involved for breach of Article 101(1)).[577]

6.1.3 Article 101 TFEU enforcement and the regulatory benefits of block exemptions regulations

Post-modernisation, the easiest way for organisations to obtain certainty in being able to avail themselves of the Article 101(3) legal exemption criteria is to fit within the terms of a variety of

[576] Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty (OJ 2003 L1/1). This is sometimes referred to as the 'Modernisation Regulation'. Different rules were previously embodied in Regulation 17/62 (OJ 1962 L13/204) pursuant to which – before 1 May 2004 – the European Commission had exclusive jurisdiction to decide whether the Article 101(3) criteria were met, by granting individual exemption decisions on a case-by-case basis. One consequence of this exclusivity was that there were large numbers of notifications of agreements to it. As national competition authorities did not have the power to apply Article 101(3), some companies used the centralised authorisation procedures to seek legal security and protect themselves from private actions before national courts. This approach – based on an individual exemption system and associated notification procedure - undermined efforts to promote the decentralisation of competition law enforcement within Europe being a lengthy and administratively-heavy process.

[577] The European Commission, national courts, and national competition authorities, share joint competence to declare that the Article 101(3) TFEU exemption criteria are satisfied or not in any particular case.

block exemption regulations issued by the European Commission.[578] Briefly, an agreement which meets the qualifying conditions set out in a block exemption can benefit from a presumption of legality via automatic exemption from the Article 101(1) prohibition.[579] Therefore, organisations are incentivised to self-assess their agreements to ensure that they can gain the benefits of being automatically block-exempted, which also provides them with greater legal certainty that they are complying with EU competition law rules.

Self-assessment of agreements involves effects-centric analyses, but may also require subsequent action to be taken by companies who are parties to an agreement to reduce the likelihood/magnitude of any resulting competitive harm associated with its implementation. Said otherwise, intended agreements can be designed but also amended to fit within the terms of a block exemption regulation to ensure regulatory 'fit' upfront, the assessment of which should be documented by companies as part of their compliance duties.[580] Effectively, mitigatory steps may be taken to increase certainty that the agreements are legal and can be evidenced as such.

To aid both these tasks, the European Commission has developed a body of guidance for companies to promote coherent application of competition rules throughout the EU. These include guidelines specifically on the application of different block exemptions, and related legal instruments offering safe harbours from legal consequences, to help companies ensure with reasonable certainty that their activities comply with EU competition law. Such guidance takes into account relevant prior decisions of the Commission, together with relevant judgements by the CJEU (and the EU General Court, formerly the European Court of First Instance).[581]

The main block exemptions under Article 101 relate to: vertical agreements; technology transfer agreements; research and development; and specialisation agreements.[582] For the purpose of the

[578] To note, block exemption regulations predated modernisation. They had already been introduced for certain types of agreement in order to reduce the burden on both companies and the Commission of making and dealing with separate applications for individual exemption. However, before modernisation, such regulations tended to be 'form-based' (i.e. the form of relevant agreements was the pivotal legal test for determining their applicability) rather than 'effects-based' (i.e. looking to the consequences of relevant agreements - through a factual assessment of their likely effects on market competition - for determining their applicability).
[579] To note, block exemption regulations are not only issued in respect of Article 101 TFEU. They are also issued in the related area of state aid law that allow the European Commission to declare specific categories of state aid compatible with the EU Treaty, including exempting them from legal requirements. However, in either case, the principle is the same: they provide an automatic exemption from illegality to anyone that meets the conditions set out in a relevant block exemption regulation in respect of particular economic activities they carry out.
[580] Some basic documentation should be available in all cases regarding assessment of how a block exemption is satisfied in individual cases. It is on the basis of such documentation that such assessments may be further evaluated and possibly contested by competition agencies if they choose to so investigate certain practices. However, the extent of documentation recorded by organisations is scalable to the complexity of the associated analysis.
[581] Also throughout the state aid modernisation process, the Commission followed a consistent approach in establishing formal guidelines containing the criteria for assessing state aid compatibility.
a[582] There are also a number of sector-specific block exemptions, notably related to motor vehicles, the insurance sector, as well as various transport-related ones.

data protection comparisons drawn in this chapter, the most important of these is the vertical agreements block exemption (VBER, set out in full in Appendix 5 below), which entered into force on 1 June 2010 and automatically exempts all types of vertical agreements that fall within its scope from the application of Article 101(1).[583] More specifically, it provides a presumption of legality for such agreements where concluded by companies with 'insignificant' market power (signified by holding shares of less than 30% of the markets in which they sell or buy the relevant goods or services). Below this safe harbour threshold, appreciable anti-competitive effects are deemed unlikely to result from the relevant agreement.[584] To note, while agreements that do not fit within block exemptions are not automatically illegal, companies must make their own assessment to determine whether they in fact appreciably restrictive of competition. If such 'non-fitting' agreements are found subsequently to be so by a competition authority, such practices are automatically void and unenforceable, and the parties to them may be subject to substantial authority-imposed fines.

To accompany the entry into force of the VBER, the European Commission also published Guidelines on Vertical Restraints.[585] This formal guidance was designed to assist companies in carrying-out self-assessment of the likely effects to flow from different types of restraints (commercial restrictions on activities) commonly found in vertical agreements in complement to consideration of whether the benefit of the VBER applies.[586] It does this by setting out the factors that the Commission considers the most important in establishing whether restraints inserted into agreements would likely result in appreciable anti-competitive harm. As the importance of each factor in contributing to likely effects is admitted to vary from case-to-case depending on the factual circumstances, it is thus intended as a methodology (akin to a decision-making framework) of effects-based analysis to be followed (with accompanying examples of different types of contextual

[583] Commission Regulation (EU) No 330/2010 of 20 April 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of vertical agreements and concerted practices. 'Vertical' means that it applies to agreements between parties active at different levels in the economic supply chain, such as agreements between a supplier and distributor, or a distributor and customer. (The previous vertical agreements block exemption (Regulation 2790/1999; OJ 1999 L336/21) entered into force on 1 June 2000).

[584] Unless the agreements contain certain so-called 'hardcore' vertical restrictions (also known as object restrictions), such as restrictions on resale prices, export bans, or other restrictions conferring absolute territorial protection.

[585] Commission Notice (2010, Guidelines on Vertical Restraints), which set out the European Commission's policy in relation to assessing vertical restrictions imposed as part of commercial practices under EU competition law. While the principles laid down in them are not formally binding and are subject to approval by the European Courts on a case-by-case basis, nevertheless they offer helpful guidance on various issues that complement the VBER. To note, unlike under data protection law, the European Commission retains a key role in determining EU competition policy in areas that also affect all types of agreement, including through the adoption of enforcement decisions in leading cases against specific companies. The Commission also adopts formal Notices, which are similar to guidelines in that they are non-binding (e.g. they have no binding effect on the EU courts), but are also highly influential because they have been adopted and published by the Commission.

[586] The Guidelines address how the general principles outlined above are to be applied in analysing the following specific vertical restraints: single branding, exclusive distribution, exclusive customer allocation, selective distribution, franchising, exclusive supply, tying, and recommended and maximum prices.

scenarios).[587] Relevant extracts from the Guidelines on Vertical Restraints containing effects-based commentary (and mentioning other features with which analogies are drawn in the next Section) are set out in Appendix 6 below.

6.2 Possible models of personal data using an evolved effects-based exemption model drawing on insights from competition law

This section considers what data protection law might learn from competition law regarding the use of legal instruments providing safe harbours, i.e. automatic presumptions of legality to certain conduct satisfying its conditions. Specifically, the question is raised, 'how might creating an EU data protection block exemption regulation – developing certain features of effects-based approach modelling - improve regulatory incentives for organisations to comply with the law?'

Per Chapter 5, the prospect of law reform under which the identificatory requirement would be removed is remote. For that reason, the suggestion to abandon the concept of personal data under the GDPR is rejected, which amounts to accepting that all information relating to (living) persons would very likely be assumed to be personal data in the future. However, a very broad definition of personal data could be restricted alongside the introduction of an evolved effects-based exemption model. Effectively, this would mean that all future processing of data relating to living people by organisation should trigger at least an obligation to assess what effects such processing is likely to have.

Viewed from a jurisdictional sense, this possibility might be conceived as akin to providing organisations processing data relating to persons with the ability to discharge the *negative* (i.e. to demonstrate that their processing relates to data that should not be considered 'personal' in the traditional sense of having the full gamut of data protection rules applying to it). Said otherwise, using an exemption-based system, data controllers could be given the opportunity to obtain an automatic presumption that they are *not* processing *personal* data (albeit they are processing data relating to living persons).[588] Thus, in contrast to the Effects-based Approach put forward and discussed in Chapter 4, under this new proposal, the burden of demonstrable justification regarding

[587] These include factors related to the relevant market environment, such as barriers to entry into the market, the maturity of the market, the nature of the product or service, the duration of the agreement, and the regulatory environment.

[588] Said otherwise, organisations that process data relating to persons will have to show, ex-ante, that harm of a magnitude that is less than appreciable is likely to flow from the data relating to persons that they process – so that they can *overturn* the presumption that they fall inside the data protection regime. Alternatively, they would need to demonstrate that they comply with data protection rules, or risk being found subject to substantial penalties by DPAs for breaching such rules, as may become increasingly normal regulatory practice under the GDPR (with its higher maximum fines levels).

the personal data concept could effectively be reversed under EU data protection law by those that can fit themselves within the exemption model introduced in Chapter 5. However, as explained below, they would still have to be able to demonstrate that the harm reasonably likely to flow from their intended processing activities are *not* in fact *appreciable*.

It is in this hypothetical legislative context that a safe harbour regulatory instrument of a type used in EU competition law might be introduced. Specifically, a new block exemption regulation could herald a new jurisdictional legal sub-framework under EU data protection law, the benefit of which could be gained if certain conditions (set out in the regulation) were met in respect of a processing activity planned on data relating to a person. Reliance upon such conditions would require a likely-effects based analysis being carried out. Those able to evidence that their carrying-out of a planned data processing activity upon data relating to a person will not reasonably likely have an appreciable effect on him/her could obtain automatic legal assurances (e.g. as explained below, that they would not be fined under data protection law in respect of such activity).

Critical to this revised model is the importance of risk-mitigating controls that organisations can use to fit within the ambit of a new block exemption. Those trying to gain the benefit of the automatic safe harbour on offer would be encouraged to take steps aimed at reducing the potential for harm to data subjects flowing from their planned processing activities. Documenting the likely risk of harm from processing activities, as well as possible mitigation of those risks, are important parts of this. It should also be seen as an iterative process, and not a single task, if processing circumstances change before the activity is carried out.

Two types of possible block exemption regulation are considered, alongside their strengths/weaknesses - including consideration of the fact that in reality, the likelihood of effects-based assessments being conducted by organisations would depend on the degree of legal/policy compulsion associated with its introduction.

6.2.1 Model 1: a proposal for a jurisdictional 'de minimis' block exemption under data protection law

Under Model 1, a Jurisdictional Block Exemption Regulation (hereafter, JBER) could be issued by the European Commission that would take any processing activity applied to data relating to persons outside the scope of EU data protection rules. The justification for this would be that no

processing of personal data would be presumed to exist where the harm arising from the activity is considered de minimis (i.e. non-appreciable).[589]

The JBER would provide an automatic safe harbour – i.e., create a presumption of legal exemption from data protection rules for those that process data relating to persons that fall within its scope – binding upon national DPAs. Said otherwise, if an organisation could demonstrate that (to the best of its belief) any processing it would carry-out in relation to data relating to persons satisfied the terms of the JBER (because – in effect – it was not likely to cause appreciable harm to the data subject), a DPA would not be able to fine the organisation for any breach of data protection law if it later decided to open an investigation - and thereafter issue an enforcement decision against it - in respect of this processing. (Remembering again that DPAs can impose substantial fines on

[589] An *alternative* option also considered (but discarded in preference for the JBER proposition) would be for the European Commission to publish a Notice regarding processing of data relating to persons with non-appreciable effects (effectively, a 'de minimis' *Notice*, and called exactly that). Alternatively, it could be published by the European Data Protection Board (EDPB), which is the more powerful successor to the WP introduced under the GDPR, able to issue binding decisions in the case of certain disputes between DPAs to promote the consistent application of data protection rules throughout the EU. Under such a Notice, organisations that hold data relating to persons would be encouraged to make their own assessment - assisted by its practical, framework guidance - for determining whether such data is (reasonably) likely to have an appreciable negative effect on such persons. Such a Notice could help reduce compliance costs for organisations (and individuals) when determining whether they are caught by data protection rules, and provide a useful point of reference for DPAs, the EU courts, and national courts. On the negative side, however, a Notice is non-legally binding, so the latter would ultimately not be bound by its terms. This could dis-incentivise organisations from following the guidance in the Notice and have a chilling effect upon the (free) flow of personal data. Notwithstanding, under EU competition law, there are analogous safe harbour instruments around jurisdictional criterion with clear regulatory benefits. For example, guidance as to whether an agreement has a (non) appreciable anti-competitive effect - and are thus deemed outside the scope of Article 101 TFEU - has been set out by the European Commission in its Notice on agreements of minor importance (rev. 2014). In particular, under this Notice, it advises that an agreement will only be caught by Article 101(1) if the nature or terms of the agreement and the parties' positions on the relevant market are significant enough for its operation to have an appreciable effect on the market. In practice, the crucial factor is that the parties to the agreement do not exceed certain market share thresholds set out in the Notice.
A separate Commission Notice (2004, Guidelines on the effect on trade concept contained in Articles 81 and 82 of the Treaty) explains the circumstances indicating when agreements are in general unlikely to be capable of appreciably affecting trade between EU countries and, for that reason, also deemed to fall outside Article 101 TFEU (ex Article 81 EC). In the Notice, the Commission states that two principles must be considered closely in determining whether this jurisdictional criteria is likely to be fulfilled. According to the Commission, the concept of 'appreciability' again incorporates a quantitative element, which can be measured both in absolute terms (turnover) or relative terms (market share). It provides an indication of how such quantification may be carried out, by setting a combinatory market share threshold and a turnover threshold, under which agreements are in principle not capable of appreciably affecting trade between MSs. Moreover, the Notice also mentions that – when determining whether agreements "may affect" EU trade - it must be possible to foresee with a sufficient degree of probability on the basis of objective factors of law or facts that the agreement or practice may have an influence, direct or indirect, actual or potential, on the pattern of trade between MSs.
While not intended to be exhaustive, the aim of both these Commission Notices is to set out the methodology for the application of appreciability jurisdictional concepts and to provide guidance on their application in frequently occurring situations to help companies carry out self-assessment of their agreements. Also, while having no binding effect on national competition authorities and national courts, who are not therefore obliged to apply them, these Notices are intended to provide highly persuasive guidance to these authorities and courts as to when Article 101 TFEU may be deemed non-applicable in respect of certain agreements. In particular, the Commission will not itself open enforcement proceedings against any company that falls within the scope of either Notice. Furthermore, where the Commission has opened competition law proceedings against a company for breaching Article 101, and yet the relevant company can demonstrate that it assumed in good faith that it met the criteria set out in one or both Notices, the Commission will not impose fines if it goes on to reach an enforcement decision against the company.

organisations under the GDPR, as they can under EU competition law).[590] Moreover, national courts could also be bound to apply the terms of the JBER in the sense that they would have the power to assess retrospectively whether a processing activity fulfils its conditions ex-post and, therefore, whether an exemption should be presumed to have been obtained successfully in any relevant cases in front of them, and its legal implications.

6.2.1.1 Insight drawn from Article 101(1) TFEU policy applied to Model 1 to improve legal incentives

Attempting to gain the benefit of the JBER safe harbour would provide an extra incentive upon data holders to do more than under the existing status quo. If an organisation does not fit within the terms of the JBER (or at least could not evidence its belief that it met its terms), it would be presumed that any processing of data relating to people it carried out was subject to data protection rules. Furthermore, organisations that de-anonymise data currently – while already making attempts to mitigate the risk of such data being re-identified - would only be more incentivised to continue to do so and at least get the benefit of the JBER.

For reliance upon the safe harbour, however, organisations would have to prepare a paper-trail to demonstrate their processing activities benefitted from the JBER – a standard practice for companies where block exemptions regulations are involved, such as in competition law (see fn.580 above). Therefore, if companies argue that they fall inside the JBER, whereas in reality they cannot evidence this belief upon request, they risk being fined for omitting to comply with the full set of data protection obligations falling on data controllers. Again, this would provide incentives to assess and minimise the likely privacy-harmful effects of their processing activities.

Moreover, per Chapter 5, such assessment and paper-trail preparation would not add anything *fundamentally* new to the duties of organisations that process data relating to people, as data controllers (actual and potential) will already be required to carry-out DPIAs under Article 35 GDPR - at least when there is the prospect of engaging in a type of processing "*likely to result in a high risk to the rights and freedoms of natural persons*".[591] Furthermore, Article 30 GDPR states that

[590] In other words, the organisation would be able to argue that the data was not personal data by putting forward evidence that appreciable effects had been deemed unlikely to flow from the processing activity under consideration (taking into account different factors assessed in relation to the relevant data circumstances). A DPA would be free to accept or reject such arguments based on its own effects-based assessment on the facts of each case. Moreover, since EU Regulations have direct effect in MSs' legal systems, they can be relied on in court. This means that MS courts would have to assess whether the conditions of any new block exemption regulation had been met on the facts of any case before them. To note, while the Commission has the power to withdraw the benefit of the block exemption (in circumstances where it finds that actual effects are incompatible with legal exemption, a withdrawal decision can only take effect from the date of the decision to withdraw (rather than retrospectively).

[591] GDPR, Article 35: "*[w]here a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of*

controllers/processors with at least 250 employees have to maintain certain records of personal data processing activities.[592] For controllers, for example, such records should include – inter alia – entries regarding processing purposes, as well as technical and organisational security measures applied.

Thus, engaging in risk-assessments and documenting such exercises as required for obtaining the benefits of block exemptions, accords with the spirit of the GDPR's accountability principle. It also aligns with the practical tools that the GDPR exhorts organisations to carry-out, to mitigate risks to ensure data privacy and promote compliance before carrying-out processing activities ('by design'). Collectively, these provisions reinforce the need to take effective steps to ensure the substantive principles/obligations of data protection law are met, and requires data holders to have the necessary internal mechanisms in place to demonstrate legal compliance to DPAs if so required.[593]

6.2.1.2 Insight drawn from Article 101(1) policy applied to Model 1 to improve legal certainty

Per Chapters 4-5, under the Effects-based Approach, organisations would be encouraged to make their own assessment of their intended processing activities to determine whether they are (reasonably) likely to have a Relevant (appreciable) Effect upon the individual to whom the data for processing in each case relates. Similarly, satisfying the terms of the JBER would require organisation intending to process data relating to persons to evidence that the processing activity is unlikely to cause appreciable harm to the data subject. While this may not seem as straightforward as demonstrating that you have a market share of less than a certain percentage (by comparison with the requirement in many competition law block exemptions), there are ways in which Model 1 might enhance the internal market dimension of data protection by reducing

---

natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks". Furthermore, the WP (in WP248, p.7) comments: "*In cases where it is not clear whether a DPIA is required, the WP29 recommends that a DPIA is carried out nonetheless as a DPIA is a useful tool to help data controllers comply with data protection law*". To note, during the GDPR negotiation process, the European Parliament (in its legislative resolution of 12 March 2014, P7_TA(2014)0212) proposed to go a step further and supplement the DPIA requirement with a preceding obligation to carry out a "*risk analysis of the potential impact of the intended data processing on the rights and freedoms of data subjects, assessing whether its processing operations are likely to present specific risks*". It then listed different processing operations likely to present 'specific risks' (Amendment 127, Article 32a). While this proposal was ultimately rejected, the WP's recommendation that organisations err on the side of caution and carry out DPIAs when they are in doubt over whether a DPIA is required, arguably, implicitly reinstates this recommendation (as a precursor to the DPIA obligation) at least to some limited extent in practice.

[592] Under Article 30(5) GDPR, that obligation also extends to organisations employing fewer than 250 persons where *"the processing it carries out is likely to result in a risk to the rights and freedoms of data subjects, the processing is not occasional, or the processing includes special categories of data as referred to in Article 9(1) or personal data relating to criminal convictions and offences referred to in Article 10".*

[593] In other words, the accountability principle require data controllers to have the necessary internal mechanisms in place to demonstrate compliance to external stakeholders, including national DPAs. This requirement to document adequate measures taken to ensure compliance should facilitate greatly the enforcement of applicable rules under the GDPR (compared with the status quo under the DPD).

fragmentation and strengthening legal certainty (relative to criticisms of the Effects-based Approach and its modelling as developed in Chapters 4-5).

Specifically, the JBER could be published alongside formal (and, therefore, highly persuasive) guidance - such as an interpretative 'Communication' issued by the European Commission[594] or the EDPB (see fn.589 above) - describing the block exemption's application to help organisations carry out self-assessment. Given the broadness of the term 'appreciable effects', this would help decision-making organisations interpret this phrase to aid the practical application of the JBER, as well as analyse possible impacts of different processing activity types on individuals' rights. Said otherwise, guidance examining typical data-processing scenarios, and the attendant risks of harms is critical. To this end, the Commission or the EDPB could provide examples of the different ways in which an individual may be harmed appreciably in those scenario contexts, alongside proactive compliance steps that can be taken to mitigate those risks of harm and evidence these.[595]

By way of simple illustration: if a company intends to create an app that would only be able to count the number of steps during a single walk by an individual, without being able to combine those data with other data from and about the same individual, and in the absence of other context-specific information about the way in which the app data are to be used (such as the processing of location data), the collected data are not likely to appreciably harm the individual being non-intimate. For example, the analytic processing activity is not one from which sensitive knowledge about that person (e.g. related to their health) can be known/inferred. Gratton (2013, pp.177-178) describes this risk-of-harm analysis at this (collection-storage / disclosure) stage as assessing the likelihood of "subjective" harm arising depending on the intimate nature of the information involved and whether it is already available in the public domain.

However, if the same data were intended to be tracked over time, combined with other (more intimate and non-public) data sources associated with the same individual, and disclosed to third parties or otherwise processed for secondary usage purposes, it may well be that subsequent processing activities should be deemed liable to cause appreciable harm to the data subject. Gratton (2013) describes the 'usage' stage of the risk-of-harm analysis as properly focused on whether "objective" harm (e.g., a financial harm, physical harm or some type of discrimination) might arise. She proposes – broadly in line with this thesis author, in addition to placing emphasis

[594] Akin to the form of interpretive European Commission Communications, or Notices, as used in EU competition law to import the fact that such documents are intended to explain in more detail high-level policy.

[595] In other words, certain processing contexts are in fact much more likely to arise than others within which data controllers will need to assess the likely effect of a type of processing activity on the person to whom it relates. It may be possible, therefore, given a particular informational context, to anticipate a particular data processing type's ability to affect negatively an individual's interests (albeit in a general way). This may require an accompanying normative taxonomy to be prepared on different data usage types.

on data sensitivity and the potential for disclosure harm – that (p.208): "*[i]f there is a negative impact (or what I refer to as an objective harm, such as a financial harm, physical harm or some type of discrimination), then the information would qualify as personal information and it would have to be "accurate" and "relevant" for the intended use".* Notwithstanding, per Chapter 4, Gratton's proposals can be distinguished from those of this thesis author as her analysis is interwoven into the identificatory requirement, although in practice at times she also dismisses the need to consider identification possibilities for carrying out the relevant assessment (e.g. ibid: *"[t]he fact that the information can or can't identify a unique individual does not need to be taken into account at the use level and may usually not need to be included in the assessment test at that point").*

Another example in a particularly 'grey area' of legal uncertainty that would benefit from guidance regards profiling activities – defined in the GDPR as "*automated processing consisting of the use of personal data to evaluate certain personal aspects relating to a nature person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements"* (Article 4(4)).[596] (To note, again, Gratton's approach to harm flowing from data combining activities is different from this thesis author's view (see p.138 above)).

Such new guidance could also set out a methodology for determining the likely presence of appreciable negative effects (i.e. when the likelihood of appreciable harm occurring may be foreseeable with a sufficient degree of probability based on objective factors), including

[596] For example, in Article 29 Working Party, Advice Paper of 13 May 2013, the WP has commented in the context of discussing Article 22 GDPR in a draft form (then Article 20 of the draft GDPR as proposed by the European Commission (COM/2012/011 final)) marking out profiling as a measure *"which produces legal effects"* on them or "*significantly affects*" them: *"[t]he new EU Data Protection Regulation should provide for clear rules on the lawfulness and on the conditions for the processing of personal data in the context of profiling on the one hand, while on the other hand leave a reasonable degree of discretion to assess the actual effects – positive and negative – and the degree of intrusiveness of a specific processing type or measures on data subjects. The Working Party therefore supports an approach … which covers profiling or measures based on it to the extent only that they significantly affect the interests, rights or freedoms of the data subject. Where profiling does not significantly affect the interests, rights or freedoms of the data subject, Article 20 does not apply and the lawfulness of processing is to be assessed in the light of the other provisions of the Regulation. However, **given the broadness of the term "significantly affect", a mechanism is needed to interpret and specify this phrase for practical application. This mechanism should not only take the scope of the basic right to data protection into account. It should also assess the interests of controllers and should comprise an analysis of possible and actual impacts of profiling technologies on data subjects' rights and freedoms. In the view of the Working Party, this task could best be performed by the European Data Protection Board, which should be empowered to issue guidelines on the interpretation and application of Article 20 in specific processing contexts**." (p.4, emphasis added). The same comments would also appear to hold true when considered in the context of the term 'appreciable effect'. Notably, in this same Advice Paper, the WP does not address the scenario when profiling activity is carried out on data relating to persons that is *not* first deemed personal data. In other words, there appears to be a presumption that data controllers must determine that data are personal data before they consider the rules on profiling set out in the GDPR. However, this leaves a potential regulatory gap, whereby some organisations might disavow the application of such rules on the basis that they are not processing personal data as no one is identified/identifiable from the data they process. This gap would only be picked up if the regulator (the DPA) brought enforcement action against the organisations and could then subsequently 'prove the positive' that the identificatory element was indeed satisfied.

personalisation for assessment of particular processing activity categories that typically arise.[597] The WP might be seen as having started a similar process of effects-based methodological development in relation to at least one category of similar processing activities associated with an activity genre (online advertising) in its fairly recent 2017 Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP251):

- It states (p.11): *"[i]n many typical cases targeted advertising does not have a significant effect on individuals, for example an advertisement for a mainstream online fashion outlet based on a simple demographic profile: 'women in the Brussels region'. However it is possible that it may do, depending upon the particular characteristics of the case, including: the intrusiveness of the profiling process; the expectations and wishes of the individuals concerned; the way the advert is delivered; or the particular vulnerabilities of the data subjects targeted*". Other examples are given in this guidance intended by the WP: to help data controllers assess effects from automated processing of personal data; to highlight the different degrees of impact that can result to individuals depending on the type of automated processing activity and case facts; as well as to provide recommendations (e.g. at p.14, that data controllers give data subjects "*real, tangible examples of the type of possible effects*" that can result, regardless of complexity. It gives the example of insurers using automated decision-making processing to set motor insurance premiums (based on monitoring customers' driving to identify bad behaviours) illustrating the significance and envisaged consequences of the processing – potentially higher premiums automatically charged to drivers due to their driving style - by providing an app "*comparing fictional drivers, including one with dangerous driving habits such as fast acceleration and last-minute braking*").

- It also refers to a separate Opinion WP244 in this guidance as providing useful examples that may be helpful when considering the effects of automated decision-making on data subjects. WP244, in turn, sets out (at p.4) how interpreting whether a personal data processing meets a standard of "substantially affect[ing]" data subjects should be carried

[597] As well as EU competition law comparisons potentially drawn here with the Guidelines on Vertical Restraints described above, another useful point of comparison are the Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to horizontal co-operation agreements (2011 OJ C11/1) which also sets out an analytical framework in terms of general principles of EU competition law for carrying out effects-based assessments under Article 101 TFEU, including acknowledgement of when agreements are unlikely to be caught by Article 101(1) due to non-appreciable effects on the market (to such to such an extent that negative market effects as to prices, output, innovation or the variety or quality of goods and services can be expected). In particular, it states that an effects analysis will require an assessment of the agreement in its economic context, taking account of the nature of the agreement, the parties' combined market power and a number of other structural factors such as the size and number of the parties' competitors and the existence of barriers to enter the market. Guidance on how such analytical frameworks are then applied practically in respect of particular categories of commercial activities are then provided in both instances in such guidelines by the European Commission (see, e.g. Appendix 6 below, at pp.310-314).

> out on a case-by-case basis, but taking into account certain methodological factors for the assessment of effects regarding: *"the context of the processing, the type of data, the purpose of the processing and factors such as whether the processing: causes, or is likely to cause, damage, loss or distress to individuals; has, or is likely to have, an actual effect in terms of limiting rights or denying an opportunity; affects, or is likely to affect individuals' health, well-being or peace of mind; affects, or is likely to affect individuals' financial or economic status or circumstances; leaves individuals open to discrimination or unfair treatment; involves the analysis of the special categories of personal or other intrusive data, particularly the personal data of children; causes, or is likely to cause individuals to change their behaviour in a significant way; has unlikely, unanticipated or unwanted consequences for individuals; creates embarrassment or other negative outcomes, including reputational damage; or involves the processing of a wide range of personal data".*

The new guidance being proposed by this thesis could make the law more comprehensible to organisations (than it is under such existing guidance, however) by introducing a level of objective rigour to effects-based assessments based on multi-tiered considerations. Not least this would be because it could take insight from existing DPIA/PIA frameworks of analyses,[598] whereas the need for authoritative guidance *and* a methodology for assessing data protection impacts is also implied by Article 35 GDPR and its reference to several factors (the "*nature, scope, context and purposes*") for consideration in determining whether a high risks to individuals' rights/freedoms is likely to result from a type of processing activity.[599] Hence, developing the theory and practice for carrying-out an effects-centric assessment would involve setting out relevant factors, and how analysis of such factors should be carried out, and their importance, in determining with what likelihood and magnitude of negative impact may flow from any particular processing activity.[600]

---

[598] See, e.g., Article 29 Working Party (2013, WP209), Article 29 Working Party (2011, WP180), and ICO (2014, #2).

[599] The WP has previously stated – in that context (WP218, /7) – that the "*severity and the likelihood of the impacts on rights and freedoms of the data subject constitute elements to take into consideration to evaluate the risks for individual's privacy"* in relation to which the GDPR *"already contains the criteria needed to assess the privacy risk posed by particular processing*" (see also comments in WP217 referred to in fn.429 above). Compare, Article 29 Working Party (Annex to Letter to Mr Paul TIMMERS Director of Sustainable and Secure Society Directorate DG Connect, Brussels, 25 February 2015) discussing possible indicators that health data are being processed and the WP's comments (p.4) that to assess privacy impact – in particular, to distinguish "*[r]aw, relatively low privacy impact personal data*" from higher impact personal data - "i*t does not suffice to look at the character of the data as is"; "intended use must also be taken into account, by itself, or in combination with other information*".

[600] By setting out the factors that are formally considered the most important in establishing whether a processing activity type would likely result in appreciable harm resulting – and while the importance of each factor in contributing to likely effects will vary from case to case depending on the facts and circumstances – it could be developed into a methodology or framework of effects-based analysis to be followed (with accompanying examples of different types of contextual scenarios). In other words, assessments based on seriousness of likely harm may be context-dependent, but within any given context, those planning to process data relating to persons can prioritise risk by asking set questions, such as 'what is the intended use?', 'what risks to the individual attend that use and how likely is it for that harm to occur?', and 'how severe is the likely harm suffered?'. By comparison, Hartzog & Rubinstein (2016) recommend seven risk variables for consideration: volume of data; sensitivity of data; recipient of data; use of data; the data subject's consent/expectations; data treatment techniques; and, data access controls. The inclusion of the last two factors

Furthermore, to distinguish proposed new guidance from impact assessment (PIA/DPIA) guidance issued already as mentioned, it could include a developed methodology for carrying-out such assessments to complement these – and also the JBER - for consideration *in toto*. Said otherwise, it would highlight the need for a broadly 'joined-up' compliance approach encompassing consideration of jurisdictional and substantive issues altogether (by-design), by recognising that analysis of impact can help identify situations where data protection concerns are liable to arise. On the other hand, the assessment of harm would be tailored to a jurisdictional context and understood in that context. In other words, it would not be interchangeable with reference to an impact assessment exercise already stipulated to be carried out in the GDPR. Whereas presenting the precise details of such methodology for use under an effects-based approach goes beyond the scope of this thesis.

Generally, therefore, following similar types of methodologies for assessing potential impact following the processing of data about people would help organisations structure their compliance reviews regarding, not just whether data protection rules should apply in the first place, but also, if so, what obligations would therefore arise. Effectively, it would present a broadly comprehensive data impact/harm assessment framework. Specifically, it would indicate the appropriateness of extending *similar* types of risk analyses to appreciation of the broader set of obligations associated with a particular processing activity (and sets of similar processing activities, such as profiling activities). Equally, it should help organisations reconsider the issue of jurisdiction when assessing their substantive data protection obligations regarding intended data processing activities. This approach would help overcome the challenge that data intended for processing may have been collected some time ago and compliance obligations associated with its processing wrongly forgotten (because it was assumed not to be personal data initially).

Such an approach would also clearly align with the GDPR's regulatory direction, which, as mentioned, refers to a number of practical measures/tools that can help mitigate privacy risks, to help sustain and reinforce consideration of impacts on individuals' rights at all key stages in a processing operation where data relating to persons are used. It is also consistent with recommendations by the WP in encouraging safeguards of a type that could help to prevent undue impact on data subjects.[601] Introducing an effects-based assessment-methodology would also help

indicate that, for these authors, they also recommend focusing on encouraging data holders to carry out appropriate processes for minimising risk as far as possible.

[601] See WP217 (p.51): *"additional safeguards to prevent undue impact on the data subjects, including: - data minimisation (e.g. strict limitations on the collection of data, or immediate deletion of data after use); - technical and organisational measures to ensure that the data cannot be used to take decisions or other actions with respect to individuals ('functional separation'); - extensive use of anonymisation techniques, aggregation of data, privacy-enhancing technologies, privacy by design, privacy and data protection impact assessments; - increased transparency, general and unconditional right to opt-out, data portability & related measures to empower data subjects."* Compare,

cement a coherent foundation to theories of effects-based harm under data protection law in terms of scoping not only when individuals' rights require protection, but also why, and how to protect such rights (either in terms of ensuring the legality of processing assessed in the light of the substantive data protection rules, or to take the processing outside the data protection regime).

To note, however, issuing guidance – in the hope of providing more legal clarity, thus strengthening practical implementation and raising awareness of the rules - is not considered a total solution to the problem of legal certainty (facilitating effective and consistent application of data protection rules pan-EU) as discussed further below. For example, the issuing of an interpretative Communication or Notice by the Commission (or EPDB) would remain non-binding and therefore would have only limited impact on reducing legal certainty and resulting costs, and thus enhancing the internal market dimension (compare fn.589).

6.2.2 Model 2: a proposal for a partial 'de minimis' block exemption under data protection law

There is a second Model that deserves careful consideration. Model 2 also assumes a shift to the default setting that all data relating to people is likely to be considered personal data in the future and proposes the introduction of a block exemption regulation into EU data protection law providing an automatic legal safe harbour. However, unlike Model 1, it is not proposed to be a definitional or jurisdictional block exemption. That is, Model 2's application would not take a processing activity outside the scope of EU data protection rules entirely (in reality, as described, such as to provide an exemption from the possibility of later fines by a data protection agency).

Rather, the application of the type of block exemption regulation (safe harbour legislative instrument) proposed under Model 2 would be one that enables different systems of scalable data protection rules to be built. It would do this by building on the possibility that automatic legal exemptions could be obtained from the requirement to carry-out *some but not all* EU data protection rules in respect of certain processing activities. In other words, it could make possible data controller exemptions from carrying-out certain data protection obligations in respect of a processing activity upon a particular piece of information (in terms of preventing the data controller

WP203 (pp.26-27): "*the implementation of additional technical and organisational measures may be particularly important. … Examples of the relevant measures may include, among other things, full or partial anonymisation, pseudonymisation, or aggregation of the data, privacy enhancing technologies, as well as other measures to ensure that the data cannot be used to take decisions or other actions with respect to individuals".*

from being later penalised by DPAs for not carrying-out data protection rules in respect of such processing activity).[602]

Like Model 1, obtaining the safe harbour available under the Model 2 block exemption depends upon whether conditions are met in respect of processing activities for carrying-out upon data relating to a person (deemed personal data in all cases under this model). Moreover, the latter's application would also be triggered through evidencing the satisfaction of such conditions indicating that any harm flowing from said processing activity was not reasonably likely to be appreciable.

The Model 2 block exemption would also be binding on national DPAs providing a legal presumption that certain data protection rules were not breached in respect of a data controller carrying-out processing activities that met its conditions.[603] Moreover, national courts would have the power to assess whether a processing activity fulfils its conditions and, therefore, a partial exemption from

[602] Exemplar partial 'de minimis' safe harbour formats that Model 2 could be modelled upon include examples arising in the area of EU state aid law, which is closely associated with EU competition law. Under Article 107 TFEU, state aid is defined as any transfer of MS's resources which creates a selective advantage for one or more business undertakings, has the potential to distort trade between the relevant business market and affects trade between the MSs. Such practice is unlawful for being generally incompatible with the aims of the Single Market. However, exemptions from this prohibition are available where state aid can be justified by reasons of general economic development (deemed to bring benefits to society that outweighs the possible distortions of competition in the EU). Briefly, typically state aid schemes are notified for approval to the European Commission pursuant to state aid rules. The European Commission then carries out an evaluation aimed at verifying whether the assumptions and conditions underlying the compatibility of an aid scheme have been complied with, whether its objectives have been realised, and the effects of the aid on competition and trade. Where schemes are not so notified, it is because EU MSs (through their public bodies) have decided that either state aid is not involved or they have assessed that exemptions may be available to them because the benefits of their aid may be deemed to outweigh any possible negative effects on competition resulting from the aid.
In this context, the Commission has made regulations in relation to categories of aid that provide aid givers with exemptions from a number of obligations including the obligation of prior notification to it under the state aid rules. In 2008, a General Block Exemption Regulation relating to state aid was introduced, which sets out categories of state aid exempted automatically from prior notification to the Commission for approval, along with the conditions under which aid measures can benefit from such exemptions. There is also a 'de minimis' Regulation, which exempts small subsidies from the notification requirement. In 2014, the European Commission adopted a revised block exemption GBER for state aid (Commission Regulation (EU) No 651/2014 of 17 June 2014 declaring certain categories of aid compatible with the internal market in application of Articles 107 and 108 of the Treaty, OJ 2014 L187/1, hereafter GBER) as part of modernising reform in this area, with similar intentions. The Commission also provides guidance to public bodies in assessing whether their proposals for aid measures are exempt because they satisfy the conditions laid down in the GBER.
In terms of the type of conditions found in the GBER, these are derived from the Commission's experience and decision-making practice, especially from the application of previously published state aid frameworks and guidelines. First, the GBER sets out quantitative thresholds (such as specific financial thresholds) below which certain types of aid are exempt from certain obligations. Second, more qualitatively, Article 5 GBER stipulates that its provisions only apply to aid that is transparent and references are made to the conditions that must be satisfied in this respect, or otherwise require an appreciability risk-assessment to be carried out. Notwithstanding, the fact that a state aid measure is not covered by the GBER does not imply that it is incompatible with EU state aid rules. It merely means that the measure needs to be notified to the Commission for assessment under existing guidelines and frameworks, or run the risk of being investigated subsequently by the Commission at the behest of a complainant for unduly distorting competition in the EU.

[603] In other words, if an organisation could demonstrate that (to the best of its belief) any processing it would carry out involving data relating to persons satisfied the terms of the block exemption, DPAs would not be able to fine the organisation for *certain types of data protection law breach* if it later decided to open an investigation and thereafter issue an enforcement decision in respect of this processing.

data protection rules could be presumed to apply in respect of such activity in any relevant cases in front of them.

In relation to which rules would be deemed exempted, these would be set out in the block exemption (discussed further below), broadly conceived as those data controller obligations which are deemed disproportionate to the risks of harm involved.

6.2.2.1 Advantages of Model 2 over Model 1

Like Model 1, the effect of Model 2 would also be to encourage organisations to take action to mitigate the risks of harm arising from a processing activity – as well as demonstrate compliance and accountability related thereto - so they can in turn avail themselves of an exemption from having to carry-out certain data protection obligations. In other words, those processing data relating to persons would have incentives to put in place measures to mitigate the risk of harm from intended processing activities to a non-appreciable level (where previously such activity would be assessed as reasonably likely to have an appreciable effect).[604]

However, the main reason for proposing Model 2 as an alternative to Model 1 is that the latter may be considered to sweep too widely, in that there is larger risk under Model 2 of false negatives occurring (i.e. processing activities incorrectly being considered to meet the safe harbour conditions under Model 1 – being adjudged non-appreciable in potential effects - wherein in fact they do not). Model 2 provides for more proportionate (modulated) protection, rather than a simple on/off (binary) switch for the application of data protection law. Said otherwise, Model 2 allows for more scalability in the application of data protection obligations - with less chance of chilling data processing activities - while also protecting individuals' rights in data. In turn, such an approach may be seen as supporting promotion of the twin policy aims of data protection law.

In terms of increasing legal certainty under Model 2, a similar approach could be taken as proposed under Model 1. Over and above a proposal for issuing guidance to accompany a new block exemption, one way to demonstrate the functionality of this new approach is by considering which data protection rules might be exempted and why. This is discussed in Section 6.3 below. For the remainder of this section, it is discussed how - while the latter issue might seem a mammoth task in its own right - there is already policy and academic literature support for a modulated approach to the non-application of certain data protection rules depending on the circumstances considered

[604] Moreover, as with competition law (albeit a rare occurrence in practice), the fact that the courts and regulators can still intervene when it is clear that a block exemption is falsely being claimed (i.e. when there is evidence to suggest that its terms could not be said to be satisfied according to the controller's best beliefs on the facts, as documented), means that companies will not to be incentivised to skew the facts upon which their self-assessments are made.

in the context of the identificatory-approach to the personal data concept. This is described next and how such support might align under the Model 2 proposal for a tiered approach that acknowledges less need for personal data protection (at least in respect of the application of certain data protection obligations) when there is a lower risk of harm to data subjects.

6.2.2.2 Support for the proportionate approach envisaged under Model 2

Flexibility is already built-into data protection rules, to some extent, for providing a risk-appropriate legal response to the processing circumstances under consideration. Per Chapter 4 - despite the broad concept of personal data contained in the DPD/DPA/GDPR - the mere fact that a certain situation may be considered as involving the processing of personal data does not alone dictate the type of data protection rules applying. Specifically, the legal concept of sensitive data shows how data protection legislation already supports the notion that different levels of data protection obligations (upon the processing of different categories of personal data) may better protect individuals' privacy and other rights in data. Said otherwise, sensitive personal data are provided with stronger legal protections (through stronger obligations upon controllers to comply with data protection law) when processed than non-sensitive data types.[605]

Analogous arguments have been used in recent legislative/policy discussions as part of proposals to develop data protection law.

6.2.2.2.1 Policy support

In the GDPR reform-discussions, the European Parliament proposed that - as data that has been pseudonymised is associated with a low levels of risk associated with its processing - it may be justifiable to exempt data controller from certain data protection obligations when processing such data. To effect this suggestion, in its legislative resolution,[606] the European Parliament recommended adding a formal definition for "pseudonymous data" to the GDPR,[607] alongside a

[605] Furthermore, even where the processing of (non-sensitive) personal data within the scope of the DPD is involved, not all the rules contained therein may be applicable in the particular case. A number of provisions of the DPD contain a substantial degree of flexibility, so as to strike the appropriate balance between protecting data subjects' rights on the one side, and on the other side facilitating the achievement of legitimate interests of data controllers, third parties, and the public interest which may be present. For example, see Articles 6 (retention period depending on data being necessary), 7(f) (balance of interest to justify processing), the last paragraph of 10 (c) and 11.1 (c) (information to the data subject where necessary to guarantee fair processing), or 18 (exemptions from notification requirements). Moreover, as reflected now in Article 22 GDPR, the WP has previously given its support for the obtaining of *explicit* consent to be necessary (when consent is the legal basis being relied upon) by those engaging in the processing of personal data who wish to carry out automated decision-making activities that have "legal/significant" effects on data subjects (see Article 29 Working Party, Advice Paper of 13 May 2013).

[606] European Parliament legislative resolution of 12 March 2014 (P7_TA(2014)0212).

[607] Ibid, in its proposal of the time for a new Article 4(2)(a), as meaning personal data "*that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution*" (Amendment 98)).

sliding-scale, risk-based approach to personal data protection obligations for such data. For example, it proposed that the restrictions upon the data controller with respect to carrying-out certain profiling activities[608] would not apply to profiling activities carried out upon pseudonymous data. This, the legislative resolution explains, is because "*profiling based solely on the processing of pseudonymous data should be presumed not to significantly affect the interests, rights or freedoms of the data subject*".[609]

The Council of the EU similarly endorsed a legislative risk-sensitive approach during the GDPR reform discussions, whereby certain data protection obligations may be 'turned off' in certain personal data processing circumstances. While the Council did not agree with the Parliament about including a new legal category of personal data ('pseudonymous data') in the GDPR, it suggested a reduction in compliance obligations with respect to the processing of data in relation to which a data controller "*is not in a position to identify the data subject*" (and can demonstrate as such if required). Specifically, Article 10 of the Council's General Approach to the GDPR states, "*if the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller*" – and the same data controller "*is not in a position to identify the data subject*" (and can demonstrate as such if required) - a number of data subject rights would become inapplicable.[610] The Council suggests these inapplicable rights should relate to a data subject's right of access (Article 15), their right to rectification (Article 16), their right to erasure / to be forgotten (Article 17), and their right to portability (Article 18).[611]

While the (finalised) GDPR does not follow the Parliament's or the Council's proposals in all respects, generally preferring milder solutions, it gives credence to the idea that applying its full requirements to the processing of identifiable data could, at times, create perverse results in obliging organisations to collect more personal data solely to authenticate data subjects. Article 11 GDPR provides that data controllers need not collect more personal information to identify the data subject for the mere purpose of complying with the GDPR.[612] The logic of the Council's approach on

[608] Ibid, Article 20, which stated that a person may be subjected to profiling that leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject, but only if certain conditions are met (as set out in its Article 20(2)).
[609] Ibid, Amendment 34, Recital 58a. In cases where the data controller bases their processing activities on the legitimate interest ground (Article 6(1)(f)) the European Parliament also included an assumption that, "*provided that the interests or the fundamental rights and freedoms of the data subject are not overriding, processing limited to pseudonymous data should be presumed to meet the reasonable expectations of the data subject based on his or her relationship with the controller*" (*Amendment 15, recital 38*).
[610] Council of the EU preparation of a general approach of 11 June 2015 (565/15).
[611] The Council also suggested the insertion of a new Article 12(4a) in the GDPR, to require data controllers to request additional information from an individual asserting a right if they have reasonable doubts about their identity.
[612] Article 11(2) provides an exemption from the rights to access, rectification, erasure and data portability outlined in Articles 15-20 GDPR if, "*the controller is able to demonstrate that it is not in a position to identify the data subject*" (except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification). Compare Article 11(2) DPD: *"[p]aragraph 1 [Information where the data*

this point is retained: differing regulatory requirements can be justified as appropriate depending on the risk-levels broadly associated with different data situations, going beyond determinations that (statically-conceived) data-type labels exist.[613]

From a regulator viewpoint, WP136 notes the "*considerable flexibility in the application of [data protection] rules*" and exhorts that, in light of this flexibility, "*[i]t is a better option not to unduly restrict the interpretation of the definition of personal data"*.[614] Also in WP187 (its 2011 opinion on geolocation data on smart mobile devices), the WP found some personal data (a Wi-Fi MAC address in combination with location information) deserved a lighter set of obligations because it posed a lesser threat to privacy ("*of the owners of these access points than the real-time tracking of the locations of smart mobile devices"*).[615]

The ICO has also set out its support for a proportionate approach to data protection obligation. In 2009, it suggested that different kinds of information, such as IP address logs, might have different data protection rules applied to them (disapplying certain obligations in respect of their

---

have not been obtained from the data subject] shall not apply where, in particular for processing for statistical purposes or for the purposes of historical or scientific research, the provision of such information proves impossible or would involve a disproportionate effort or if recording or disclosure is expressly laid down by law. In these cases Member States shall provide appropriate safeguards."

[613] See, e.g. Schwartz & Solove (2013, #1, pp.913-4) discussing this provision against the backdrop of assessing levels of re-identification risk: *"identifiable information should not and cannot be regulated in the same manner as identified information*. *Thus, while the Proposed Regulation does not specifically create two classes of personal data with differing requirements, Article 10 would permit such results*". Schwartz and Solove go on to say, *"while Article 10* [now Article 11 GDPR] *recognizes that identified personal information should be regulated differently from identifiable information, it is an incomplete solution… [to address] this need for nuance regarding classification of personal information".*

[614] WP 136, p.5.

[615] WP 187, p.16. In discussing this WP quote, Schwartz & Solove (2013, #1, p.914) comment: "*The Working Party's ultimate conclusions about Wi-Fi routers demonstrated a modest, initial step that may lead, one day, to evolution of the EU's view toward PII. It called for a less rigorous opt-out mechanism, rather than an automatic opt-in, as well as a lighter notice requirement, and it implied that access for the affected individual need not be provided if provision would require collection of additional information to authenticate the Wi-Fi access point owner. Initially, the Working Party broadly stated that a data controller should treat "all data about WiFi routers as personal data." Even when "in some cases the owner of the device currently cannot be identified without unreasonable effort," a Wi-Fi access point should be viewed as personal data. It reached this conclusion because the information can be indirectly identified in certain cases. Thus, the opinion of the Article 29 Working Party did not demonstrate flexibility in the definition of "personal data." Its starting point was that a Wi-Fi MAC address in combination with location information constituted "personal data." Yet, it also found that this information posed a "lesser threat to the privacy of the owners of these access points than the real-time tracking of the locations of smart mobile devices"".*

For completeness, set against this interpretation of the WP's approach, however, is a statement by the WP in the Annex to Letter to Mr Paul TIMMERS Director of Sustainable and Secure Society Directorate DG Connect, Brussels, 25 February 2015, p.7: "*[t]he Working Party recalls its Statement on the role of a risk-based approach in data protection legal frameworks of 30 May 2014, where it has underlined that fundamental principles applicable to the controllers (i.e. legitimacy, data minimization, purpose limitation, transparency, data integrity, data accuracy) should remain the same, whatever the processing and the risks for the data subjects. The Working Party simultaneously expressed its concern about the introduction of the notion of a lighter data protection regime for pseudonymised data. While pseudonymisation can represent an important safeguard with regard to for example data security, the use of pseudonymous or pseudonymised data is, in itself, not sufficient to justify a lighter regime on accountability obligations".* Compare also a comment in a speech by EU Commissioner Reding (The EU data protection Regulation: Promoting technological innovation and safeguarding citizens' rights, 4 March 2014, [online]), *"pseudonymous data must not become a Trojan horse at the heart of the Regulation, allowing the non-application of its provisions",* in the context of concern that the introduction of a pseudonymous data category might lead to a level of protection that is too low.

processing).[616] In 2011, it commented more generally that, *'a simple "all or nothing" binary approach to the application of data protection requirements no longer suffices, given the breadth of information now falling within the definition of "personal data"'.*[617] Finally, in the context of GDPR reform discussions, the ICO stated, *"[w]e prefer a wide definition of personal data, including pseudonymised data, provided the rules of data protection are applied realistically, for example security requirements but not subject access...."*[618]

6.2.2.2.2 Academic support

Some scholars advocate the introduction of a modulated application to data protection obligations when differing risk levels are associated with different types of data processing situations, thereby introducing more proportionality into the legal framework. Specifically, some argue that the law should recognise the need for some intermediate category of data being subject to the application of some but not all data protection rules when processed.

Although this support is advocated in the context of the identificatory-approach to the personal data concept, comparisons with Model 2 may be read-across, especially as some scholars refer specifically to risks of harm flowing from processing activities as relevant considerations. For example:

- Tene suggests that "*the nature of data as personal or not could be viewed as a continuum, as opposed to the current dichotomy. This means that data which are only identifiable at great cost would remain within the legal framework, yet be subject to only a subset of fair information principles".*[619]

- Hon et al propose an *"accountability-based approach to address the privacy concerns originally intended to be dealt with by the DPD 'personal data' concept".*[620] This, they suggest, would entail *"moving to a more nuanced, proportionate and flexible regime"*, which bases *"the applicability of data protection obligations on the risk of harm and its likely severity, with appropriate exemptions".*[621] Specifically, they suggest a two-stage approach: "*[f]irst, the definition of 'personal data' should be based on the realistic likelihood of identification. Secondly, rather than applying all the Principles to information which has been determined to be 'personal data', there should be consideration in each particular*

[616] Such as, in the case of IP logs, requiring security measures, but neither the availability of subject access rights about them, nor consent to log recording (IC, 2009, p.2).
[617] See also, e.g., IC (2010, p.7) and IC (2011, p.8).
[618] ICO (2013, pp. 6-7).
[619] Tene (2011).
[620] Hon et al (2011, p.44).
[621] Ibid, p.46.

context of which data protection rules should be applied, and to what extent, based on the realistic risk of harm and its likely severity…. the risk of harm and its likely severity should then be assessed, and appropriate measures taken in relation to the personal data, with the obligations being proportionate to the risks".[622]

- US scholars Solove/Schwartz propose a categorisation scheme, titled "PII 2.0", to replace the binary 'PII/non-PII' distinction (across US and EU law).[623] Inherent to this risk-based scheme is a distinction between 'nominally identifiable information' ("*in which linkage to a specific person has not yet been made, but where such a connection is likely*"[624]), and other types of identifiable information. They argue, while nominally-identifiable data should be treated equivalently to identified data and subject to the full gamut of privacy and data protection rules, this is not true of other types of identifiable information from which specific identification of an individual, while possible, is not a significantly probable event. They argue the latter should fall within an intermediary category of data in the middle of their risk continuum, which they propose should be subject to some but not all privacy restrictions.[625] In turn, they suggest that regulatory obligations (by way of required 'fair information principles' (FIPs) – the US equivalent to data protection obligations - imposed on the controller) should be tailored to this assessment and even go so far as to contrast their approach with "*the EU approach to PII which defines it quite broadly and provides a full suite of rigorous protections to a wide array of data* ***even when there is little risk of harm***" (emphasis added).[626] The authors perceive the risk level associated with intermediate category data as low to moderate, which justifies information of this type being regulated differently. Moreover, they acknowledge that a data controller may put

[622] Ibid, p.44. To distinguish Model 2 from this approach it is not being proposed here that consideration of which rules should be (dis-)applied should be considered in each context, rather than this should be pre-set within the terms of the block exemption regulation. Moreover, Model 2 would apply, not just in relation to data that has been subject to anonymisation techniques but more widely to any data relating to persons and it would not depend upon an initial identificatory analysis (i.e., contrast, according to the two-stage analysis of Hon et al (2011, p.44), their second stage of analysis would apply only if, "*the resulting risk of identification is still deemed sufficiently high [such that] the data be considered 'personal data', triggering data protection obligations; and, appropriate technical and organisational measures should be taken to minimise the risk of identification"*.

[623] Schwartz & Solove (2011, p.1814). See also Schwartz & Solove (2013, #1, p.877).

[624] Schwartz & Solove (2013, #1, p.907).

[625] Schwartz and Solove (ibid) give the following example: "*[a]n example of identifiable information under the PII 2.0 model would be the key-coded medical data that the Working Party discussed in its "Opinion on the Concept of Personal Data". Some or all of this information might never be identified. Depending on the risk scenario, there may be only a remote chance of future linkage to a specific person. As a further example, Kuner's discussion of [an example used by Kuner of a singled out but nameless individual] the Verdi-loving physician may represent merely identifiable information under PII 2.0. Kuner's hypothetical leaves much open regarding the "data controller." We know only that the data controller himself cannot identify the person to whom the information relates. If the data controller also has strong measures in place to protect the data from exposure to others, the PII 2.0 model would classify the information as identifiable, but not identified".*

[626] Schwartz & Solove (2013, #2, p.4).

measures in place around nominally-identifiable data that lead to its classification changing so that it fell subsequently within the intermediate category.

Such support for a modulated approach to the non-application of certain data protection rules – albeit that that is does not represent the conventional wisdom – suggests an evolution of EU data protection law under which the full requirements of EU data protection law need not apply to all types of personal data depending on the identificatory-capabilities associated with the data at issue.[627] Yet – while for some this 'turning-pivot' is conceived statically such as through a new legal category of 'pseudonymous data' or 'identifiable information' that reifies a state of data that has been subject to certain processes or manifests certain characteristics – it is underpinned by an appreciation that ultimately a distinction in treatment is justified because of the likely risks that may be assessed to accompany different data environments associated with the processing of personal data.[628]

This rationale leads the way for (the more radical than Model 1) Model 2, also advocating a proportionately-tiered approach (justifying less need for personal data protection when there is a lower risk to data subjects), albeit that it focuses primarily on different levels of harm that likely might flow from a particular processing activity to data subjects. Yet, Model 2 also provides more incentives for compliance by assuming that all data relating to persons are personal data in the first place, and encouraging data controllers to tailor their plans for processing such data to fit within a block exemption by assessing the risks of harm to the data subject in the circumstances.

The theories of three other scholars should be re-mentioned here, for completeness, as their ideas are directly relevant to this proposal. First, Hartzog (2014) – as mentioned in Chapter 4 - argues that privacy protections do not always need to be substantial or direct (such as by wholesale or vigorous 'lock-down' changes in the law, which are not always suitable or cheap). Relatively modest privacy protections of diverse kinds can be effective as piecemeal remedies to particular problems individually, and even more so as a collective force. The value of such modest protections are often overlooked in a way that is misguided, argues Hartzog, in particular regarding the benefits of their flexibility in addressing different types of privacy harm. In this context, the thesis-proposed (exemption-modulated) EU regulatory response that can provide less or more protection - depending on the intended data processing context and the nature of the likely effects likely to

[627] This is despite the fact that all the authors acknowledge the importance of harm assessment to their classifications of data's nature in this context. Even Solove and Schwartz acknowledge under PII 2.0 that re-identification risk should be considered in the context of the particular analytics activity that is planned to take place. This is because, for them, PII would be identified on a risk matrix taking into account the risk, intent, and potential *consequences* of re-identification.

[628] Knight & Stalla-Bourdillon (2016, p.310).

result, in balance with other values - may yet be seen as an a nuanced and yet effective regulatory strategy. Additionally, it would help data protection be better conceptualised as an ongoing process of compliance cooperation, rather than a static framework of governance.

Second, it is worth also re-mentioning the article (2013) by Hartzog/Stutzman, which discusses a legal concept of obscurity (in particular, online obscurity) as an example of a type of modest protective measure that can be used in respect of the safe disclosure of information that does not have to remain confidential to keep individuals from harm. Obvious comparisons can be made here with the argument that indirect protections meant to reduce the probability of privacy harm rather than directly prohibit it – which apply in the context of socially-useful data sharing where hard rule-making is not always easy to lay down – is precisely the kind of nuanced legal response necessary in an age of big data analytics. For example, a Research BER (suggested in the last chapter) would enable an EU regulator to better focus on specific contextual problems related to big data analytics done for research reasons as a helpful guide/'roadmap' correctly framed, rather than trying to address too many privacy issues in one action. Both parties considering information sharing arrangements in this area would potentially be able to benefit in planning their activities (by carrying out case-by-case analysis of factors to determine whether they benefit from the BER, in a way that is good enough for most contexts); moreover, they could avail themselves of more legal certainty as part of a compliance-protective remedy as a means to minimise the risk that harm will follow without guaranteeing it. By identifying beforehand specific criteria critical to obtaining the benefit of a BER, for example, adherents would also have a clearer picture of the practices that would breach their agreement.

Third, as also reviewed in Chapter 4, Ambrose (2012) focuses on a transient and dynamic conception of data as having changing values over time dependent upon the use contexts (related to different types of information needs and the purposes for which data are collected and used, etc., for which relevance may be lost). Her exploration of this fact is in the context of identifying specific time-critical points in data processing where enforcing data protection rights (in this case, the right to erasure) may be considered reasonable (or necessary). Whereas this suggestion for a modulated approach is founded on efforts to alleviate harms caused by data accessibility, she begins to create a taxonomy to help assess competing interests at stake in different time periods in different contexts. Again, this analysis suggests the value of recognising legal mechanisms (and underlying methodologies to provide consistency) that can be relied upon when used appropriately to provide adequate privacy, while keeping the value of the processing intact. According to both theories, facilitating beneficial outcomes from processing data is a valuable aim to which regulation may be tailored without descending into a strict cost/benefit analysis balance-off. To this end, strictly binary distinctions in law about the type of data and the legal responses are rejected, as

should assessments of data processing focused on 'anonymise and forget' (each processing activity should be effects-assessed in the context of its own time point, subject to adaption if the context changes).

6.3 Evolved case study – potential application of Model 2 practically

This section aims to illustrate how the proposed Model 2 would work in practice, by applying its approach to the set of facts based on the Decision as revised in Chapter 5. First, some brief background to set the scene.

In reminder, the availability of privately-held data in the wider interest should be encouraged. For example, sharing data on the outbreak of epidemics across countries could contribute to a more timely response by medical authorities; sharing of and access to personal health data could improve diagnosis and treatment; sharing data from cars and transport means could improve traffic management and reduce congestion. Yet, facilitating data sharing (data supply and (re-)use) is often challenging. In particular, data about people might not always be accessible as a result of concerns over data privacy or perceived risks to companies' commercial interests.

The GDPR does not change this and indeed, the risk of huge fines for non-compliance – and its very broad definition of personal data, which raises concerns about when GDPR apply - may increase risk aversion by companies to sharing datasets about people. For example, in the context of the discussion in Chapter 5, the case study facts had been revised to highlight a GDPR gap in a data sharing context involving secondary analytics on pseudonymised data for research purposes. By comparison, researchers face the challenge of using vast amounts of machine-generated and semi-structured health data (e.g. of measurements, medical images, symptom descriptions) stored in large databases with the potential to boost healthcare research and innovation while ensuring that these data are processed in a secure way. It was argued that the GDPR could do more to encourage the sharing of personal data for secondary processing that may give rise to collective benefits (e.g. through data-driven innovation), while simultaneously incentivising the implementation of appropriate safeguards by organisations to reduce the likelihood of negative effects in the context of particular processing activities.

The introduction of an effects-based compliance tool oriented around safe-harbour exemptions could bolster the ability for law and regulatory process to adapt to a fast moving technology industry in a more modern, flexible, creative, and innovative way. A block exemption based on Model 2 could be such a tool. It would give data controllers the opportunity to obtain an automatic exemption from certain data protection rules providing enhanced legal comfort (comparative to an automatic presumption of legitimacy for their processing activities in many respects). Underpinning

it, and as part of the same narrative, the focus on setting up data governance structures (i.e. the arrangement of processes governing and monitoring the way data is dealt with by and between entities) should make the tailoring of data protection in view of likely effects - assessed and managed in context over consecutive time period - more feasible.

To flesh out the model, further discussion is required around: (1) how potential processing collective effects (notably benefits of a genuine societal/economic kind in the public interest) may be taken into account in determining whether a block exemption may apply; and, (2) how would organisations be incentivised to do more data sharing practically (compared with the situation under the GDPR where group harms and benefits are not specifically considered). The issues are dealt with, for reader ease, in reverse order.

6.3.1 How would the modulated approach of Model 2 apply in this case scenario?

The chosen context (per Chapter 5) again is one in which a data provider is considering sharing pseudonymised data with a private organisation, which would act as another controller once the data would be shared with it, to do generalised analytics on it for innovative purposes (i.e. to create improved output, including products and/or services). As mentioned, data analytic practices are not necessarily be considered tantamount to scientific research and doubts exist over whether they can benefit from existing research exemptions. For example, Recital 159 GDPR seems to imply that publication or disclosure of research data output should be expected to follow and Recital 33 GDPR refers to 'ethical standards for scientific research'. However, the carrying out of many private research data analytics practices will not necessarily meet these criteria.

In general, when considering data sharing, the data discloser and the potential recipient should identify upfront the objectives for the share, and then assess the legal risks depending on the specific circumstances of the data and the situation. For controllers to determine whether they could rely on the block exemption, they would also need to conduct a risk-based impact assessment, followed by the putting in place of appropriate safeguarding measures to mitigate negative effect risks to a de minimis level. The block exemption could contain specific provisions where these conditions are met exempting the application of certain data protection rules.

For example, one suggestion tailored to this case study is that the block exemption could offer an exemption from the 'purpose limitation' principle (Article 5 (1) (b) GDPR). In other words, the acts of sharing and post-sharing secondary processing could be considered automatically legitimised (*and* compatible with the initial processing purposes) *without* the need to rely upon *new* lawful bases to justify them. This exemption would be particularly pertinent in circumstances involving

data analytic practices for research purposes not carried out under the public task lawful basis (such as when carried out by universities).

More specifically, this type of lawful basis exemption would overcome some of the challenges for those wanting to secondarily process machine-generated semi-structured data, where it is not often clear whether the data are personal data (subject to data protection rules) or not. Such data processing – of a kind likely to increase rapidly in the future with the IoT - cannot easily rely on data subject consent as the exchange of data often takes place without individuals knowing about it, or may be too voluminous and fast-moving for them to understand what will happen to their data and to meaningfully agree to it. Likewise, the legitimate interest lawful basis test (Article 6(1)(f) GDPR) and the compatibility test (Article 6(4) GDPR), are plagued with interpretational problems, which are accentuated in this context. On the other hand, a block exemption could help in these situations by providing legal comfort that processing would be considered automatically legitimate: it would automatically validate that compliance that is proportionate tailored to the risks of harm to the data subject and any relevant others (e.g. groups potentially affected). In this case, upon proof of compliance with the block exemption, there would be a presumption of the legitimacy of the processing of the de-identified data. This should make collecting and sharing data about people easier, especially where complex processing and technological applications are involved.

A similar line of argument has been made by Balboni et al (mentioned in Chapter 4).[629] In reminder, they argue that data protection law must find an 'appropriate balance' in line with societal expectations to accomplish its aim of seeking to reconcile data subjects' rights and safeguards, and the free processing/flow of information across the EU. Appropriateness in this context, say the authors, can be ascertained by verifying whether the proposed processing is proportionate, i.e. delivers "*a fair balance between the demands of the general interest of the community and the requirements of the protection of the individual's fundamental rights"* (p.246). To this end, they propose that processing can achieve this appropriate balance de facto when GDPR is complied with - even if consent to processing has not been obtained, or the legitimate interest lawful basis cannot be satisfied, and no other lawful basis is relevant - because it is comprised in the corpus of its main requirements. They describe this set of obligations as forming effectively a comprehensive "Data Protection Compliance Program". Notably, however, the authors do not take into account effects-centric analysis, nor do they take into account collective interests (e.g. potential for group harm and benefits) explicitly via a modulated approach to the extent of regulatory obligations that might apply.

[629] Balboni et al (2013).

6.3.2 How would collective benefits be assessed under Model 2?

Currently, data protection law (including under the GDPR) recognise positive effects implicitly, such as through the implicit balancing of values recognised in the exceptions of Art. 17(3) GDPR that allow for retention of data: (a) to protect the right of freedom of expression; (b) for reasons of public interest in the area of public health; (c) for historical, statistical and scientific research purposes (as discussed); and, (d) for compliance with a legal obligation to retain the personal data under EU or MS law.

Model 2 would allow explicit consideration (in a legal context) of the benefits that could flow if the data sharing objectives were achieved. In cases of minimal risk, certain automatic exemptions may be obtained, conditional upon collective benefits being substantial enough to justify the existence of the harm likely to flow in the processing context. Thus, certain traditional rule-restrictions on information use and disclosure would be acknowledged as disproportionate to the residual risks associated with the usage/disclosure in practice. This would be precisely because the former would likely cripple publicly beneficial processing outcomes (e.g. socially productive uses of analytics) not raising significant risks of privacy harms.

As discussed later in this chapter, there may be detected similarities here between Model 2 and the legitimate interest test (Article 6(1)(f) GDPR). Both encourage consideration of effects/impact. However, the latter is specifically focused on cost-benefit balancing considerations that can result in very subjective findings; specifically, it allows an assessment and decision by the data controller on balancing the right to use or disclose data against data subjects' data protection rights. As discussed later in this chapter, the two frameworks can and should be distinguished. Per Balboni et al (emphasis added), "The *key issue regarding LIDC [the legitimate interests of the data controller] is the uncertainty of its interpretation, mainly due to a judgment by data controllers which remains basically subjective* ***and is rarely supported by objective standards. The lack of clear, predetermined, and objective rules impinge on the practicality of this option****"* (p.253).

Furthermore, going one step further than the concept of a Research BER outlined in Chapter 5, a different – more general – type of block exemption (hereafter 'General BER') could be created to apply to any data controllers doing data sharing/processing where collective benefits are expected and can be assessed. This proposition is not intended to preclude the additional possibility that such a General BER could be restricted in availability to certain types of data sharing scenarios (e.g. sharing with start-ups and SMEs conducive of data-driven innovation) if so desired to encourage certain types of benefits to arise as considered most desirable by regulators/law-makers. For example, DPAs could acquire the power, via delegated acts, to specify the requirements on block exemptions for different types of data processing situations.

Inspiration for such a General BER again comes from competition law, which distils four cumulative factors the satisfaction of which can be taken as indicators that a collaborative agreement/coordinated practice between organisations may generate objective economic benefits that outweigh non-appreciable negative effects of the restriction of competition. In particular:

- It must contribute to **improving** the production or distribution of goods or contribute to promoting technical or economic progress,
- Consumers must receive a **fair share** of the resulting benefits,
- The **restrictions must be indispensable** to the attainment of these objectives, and
- The agreement must not afford the parties the **possibility of eliminating competition** in respect of a substantial part of the products in question.

Article 101(3) TFEU acknowledges that, and exempts those agreements that can satisfy those conditions from the prohibition of Article 101(1) TFEU. Indeed, Article 101(3) is the legal bedrock upon which categories of agreements and concerted practices are provided an automatic safe harbour exemption (from such prohibition) under the suite of block exemption regulations on offer, assuming the organisations involved can satisfy their embedded compliance conditions. When an agreement is covered by a block exemption the parties to the restrictive agreement are relieved of their burden of showing that their individual agreement satisfies each of the conditions of Article 101(3). They only have to prove that the restrictive agreement benefits from a block exemption. These include block exemptions for beneficial collaborations (where neither organisation involved would the technical resources to produce the beneficial outputs independently) to combine complementary skills and assets and potentially leading to a wider dissemination of knowledge triggering further innovation. Examples are block exemptions for research and development and specialisation agreements, or technology transfer agreements, among others. They reflect the fact the Commission is keen to encourage collaborative research in areas of high technology so that businesses in the EU will be able to compete effectively with large international businesses that operate in markets for sophisticated/advanced products and services.

Similarly, complementarities may arise from data sharing in various ways that can be harnessed by incentivising especially innovative uses of data while ensuring that such data can be transferred in mutually beneficial ways. A block exemption is needed to guide on the how to avoid appreciably harmful ways to use data, while also ensuring that important potential benefits are not foregone in the process. As the potential benefits of new data-driven innovations are revealed, there is a need for increasing incentives for data (including data about people) to be linked across sectors and applications.

Thus, taking inspiration from competition law, four cumulative conditions could be framed, the demonstrable satisfaction of which would automatically make available the benefits of the General BER under data protection law. For example, these could ask four questions related to the sharing/processing involving data about individuals:

1. **Does it contribute to improving progress in or otherwise benefit society, i.e. it is likely to have significant public value?** The purpose of the first condition would be to delimit the types of objective benefits that can be taken into account, which must likely be clear, substantive and specific (i.e. produced only by sharing data). In effect, all benefit claims must be verifiable regarding: (a) the nature of the claimed benefits; (b) the link between the collaboration and the benefits; (c) the likelihood and magnitude of each claimed benefit; and (d) how and when each claimed benefit would be achieved.

2. **Will the arrangement also allow the relevant individuals or the general public a fair share of the resulting benefits, and can these be communicated as such to them?** This criterion would require assessment of the likely longer-term benefits and the extent to which they are likely to be passed on and the likely time lag. A 'fair share' could be determined by looking at both *breadth* (the amount of people benefitting) and *depth* (the extent to which they benefit) – or a combination of both, again based upon qualitative and/or quantitative evidence.

3. **Will the arrangement only involve using data about people that are indispensable to the attainment of these objectives?** Per the principle of data minimisation, data usage should not go beyond what is necessary to achieve the benefits to be generated by the sharing. It will therefore generally be necessary for the parties to show that the data usage are indispensable to the co-operation.

4. **Will the arrangement not involve automated decision-making befalling the individuals (data subjects) concerned in the future?** In other words, where the sharing is for secondary data-driven general analysis to be carried out by another organisation, for the General BER to be available, the data must not be processed to make decisions relating to the particular individuals whose data are re-used. The exception could be if the individuals was made aware of the information use and specifically agreed to it. Other 'black-listed' conditions could be added here (mirroring the approach in competition law block exemptions excluding so-called 'hardcore' practices – innately of a kind likely to result in appreciable negative effects - from its coverage).

These factors all accord with the spirit of data protection under the GDPR but provide a potential legal framework extending its remit of consideration explicitly to the wider positive impact of processing. Their exact formulation could be developed with further research, but would specifically fall short of requiring an explicit cost-benefit consideration of the likely benefits resulting from a particular processing activity outweighing any harm that might flow from the same activity (and if assessed outweighed, legitimacy for such activity being automatically presumed). Rather they would be embedded conditions in a General BER that must be satisfied - and demonstrable as such – separate from (over and above) satisfaction of conditions requiring risk of harm mitigation to a de minimis level. In effect, the analysis of the benefits of a processing operation would, to a large extent, be a question of evaluating whether the General BER could be availed of in the processing situation, with the four factors (related to the objective positive effects likely to ensue) setting out the cumulative criteria that must be demonstrably capable of satisfaction on the facts for this to be the case. Said otherwise, the application of the four criteria cumulatively to categories of data sharing would be presumptive of the sharing being worthy of obtaining safe harbour exemptions – comfort from enforcement action - if relevant conditions in the block exemption can be met. (If in an individual case the sharing is deemed not to fulfil the criteria, or the conditions, the block exemption may be withdrawn by a DPA).

Alternatively, as mentioned above, the availability of the General BER could be restricted to certain data sharing scenarios (e.g. sharing with start-ups and SMEs for generalised analytic (algorithm generating, after correlation are found in big data sets) research, conducive of data-driven innovation, if so desired to encourage only certain types of benefits to arise, as considered most desirable by regulators/law-makers. Furthermore, the General BER may only operate in particularly challenging areas of data protection where there is genuine uncertainty about what compliance looks like, from which public guidance and resources on compliance may be developed to lay the groundwork for future innovation.

Another alternative is that DPAs could acquire the power, via delegated acts, to specify the domains of block exemptions applied *de facto* to different types of data processing situations associated with societal benefits (i.e. without more required by the organisations involved in being able to demonstrate that the four criteria are met on the facts, it being presumed they are when certain types of sharing circumstances arise). This suggestion to create different requirements derives from the recognition elsewhere that sharing/(re-)use of data in different processing circumstances creates different forms of benefit and risk, and involves different types of actors. Attempting to govern them in identical ways would be a mistake.

Of course, the above proposed framework (a General BER or more specialised domain-specific types of block exemptions) would not be suitable for application in all data exchanges. The purpose, rather, is to provide another tool to support the sharing of personal data – especially in cases where otherwise it is not clear whether data protection rules apply in the first place – to facilitate secondary re-usage of such data for beneficial impact that can be demonstrated as such. That tool should also help remove potentially unnecessary regulatory barriers to innovation by adjusting the regulatory framework to add flexibility. Moreover, and perhaps most importantly to encourage data sharing, the General BER could be relied upon not just by the discloser of data but also the recipient in exempting them from liability in the future with respect to their secondary data processing activities. Said otherwise, both data disclosers and data recipients could get automatic benefit from the specific exemptions available under the General BER, where they can ensure (and demonstrate if challenged) that they will satisfy its conditions.

Per the accountability principle under the GDPR, it would be up to these organisations to demonstrate appropriate compliance and gain the legal comfort on offer. If they can do this, the making available of the General BER would incentivise compliance and provide flexibility by offering a standardised set of exemptions, depending on the data context, being effects-sensitive, to promote proportionate and public-beneficial outcomes founded in good data governance over time. To cite Balboni et al (p.261), quoting in their first sentence from WP173 (p.17), "*Because accountability puts an emphasis on certain outcomes to be achieved in terms of good data protection governance, it is said to be result-focused; its emphasis is on 'ex post' (ie after the data processing has started).' However, data controllers may wish to emphasize the strength of their commitment to accountability by delegating periodic … checks to independent auditors or by leveraging certification mechanisms, and data protection seals and marks".*

Next, consideration is given to some of the outstanding concerns requiring address to facilitate a move towards the Model 2 proposal.

6.4 Outstanding challenges associated with Model 2

This section examines some challenges associated with Model 2 as the thesis-preferred option for best reconciling the twin objectives of EU data protection law in developing an effects-based exemption approach in response to the inevitable broadening of the personal data concept using the (re-)identification standard and, therefore, ambit of data protection law under the GDPR. Specifically, it is clear that addressing some of the deficiencies hinted at in Chapters 4 and 5, mirrored in respect of some of the criticisms of Model 1 shifting towards the formulation of Model 2, present other key conceptual problems and research gaps. For example, more detail is required

about the conditions of the Model 2 block exemption (the meeting of which indicates that the harm flowing to a data subject from a particular processing activity is likely deemed non-appreciable). Moreover, what data protection rules might be exempted within the safe harbour provided by such a new regulation? While the case scenario discussed in the last section make some suggestions around the latter, in particular, these and other related practical and theoretical challenges associated with Model 2 are considered next.

6.4.1 Legal implementation and precedent

A practical challenge of implementation is whether introduction of a new block exemption(s) would be compatible with the GDPR. Said otherwise, would the GDPR have to be amended to allow a carve-out of its rules by another directly-applicable (block exemption) regulation? Furthermore, how would the latter be introduced as part of legislative procedure more generally?[630] These are issues for separate research not yet fully explored.

Suffice to say, inspiration for compatibility between Model 2 and the GDPR could be drawn from existing safe harbour models that currently exist under EU and UK law. For example, as mentioned in the last section, the competition law block exemptions are implemented via the introduction of secondary legislation – regulations – such as emanating from the Commission using a power to adopt and publish block exemptions. At national level, in turn, block exemptions are typically implemented by inclusion in orders made by the relevant national regulator relative to domestic law. Hence, a template for block exemption compatibility with primary legislation such as the GDPR already exists and is a 'tried and tested' approach that could readily be adapted under data protection law. Thus, Model 2 could be introduced as a carve-out of GDPR and without too much disruption, with some adaption possible at UK level via the ICO's powers.

The closest existing analogy to Model 2 in the field of data protection/privacy law is found within US federal legislation,[631] where we can find the contours of a legal class of personal information in

[630] Raising, in turn, issues about possible Commission empowerments (extensive powers to adopt further delegated or implementing acts) in the future.

[631] To note, for completeness, safe harbour principles were agreed between the European Commission and the US government in 2000 (under Commission Decision 2000/520/EC on the adequacy of the protection provided by the safe harbour privacy principles), offering a mechanism for US companies to self-certify that they provided EU-adequate lawful protection for personal data transfers from the EU to them in the US. However, in October 2015 the CJEU declared that Decision invalid (in Case C-362/14, Schrems v Irish Data Protection Commissioner [2015] 2015 ECR I-650) thereby invalidating this safe harbour arrangement. In July 2016, the European Commission adopted a superseding decision on a new framework for transatlantic data flows: the EU-US Privacy Shield. It is also based on a system of self-certification by which US organisations commit to a set of data protection principles that assume that personal data is already established as existing and planned for transfer. Said otherwise, it has the effect that transfers from a controller or processor in the EU to organisations in the US that have self-certified their adherence to the principles with the US Department of Commerce and have committed to comply with them are allowed. In both cases, however, to distinguish them from the types of safe harbour under discussion in this thesis, they relate to adequacy decisions regarding

the health sector based upon a risk-based standard reliant upon the concept of reasonableness and a methodology for its attainment.[632] The Health Insurance Portability and Accountability Act (HIPAA)[633] regulates the use and disclosure of protected health information (PHI) in any medium, as well as the collection, use, maintenance, or transmission, of electronic PHI. Specifically, the HIPAA Privacy Rule (issued by the US Department of Health and Human Services to implement HIPAA) establishes a standard for the de-identification of PHI: *"[h]ealth information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information".*[634] Those holding data that meets this standard are not subject to the legal obligations under the HIPAA.

The HIPAA Privacy Rule's de-identification standard is interpreted as some form of acceptable, low, or very small risk of linking de-identified data back to the identity of the individuals to which they correspond.[635] Two methods are described by which the standard can be met. These provide alternative 'routes' for obtaining a legal safe harbour that provides certainty for data holders when followed that they will not be subject to HIPAA's legal rules. The Privacy Rule's two de-identification methods are:

1) **The Expert Determination Method (§164.514(b)(1)).** This requires a formal determination by a qualified expert of an acceptable ("*very small*") risk based on the ability of an anticipated recipient to identify an individual; [636] or

---

substantive obligations, where it has already been established that personal data exists, rather than in the alternate 'jurisdictional' sense (as described – to be understood in a scoping sense of whether legal rules apply in the first place- in this thesis) regarding whether personal data may be said to exist or not and, therefore, whether the law applies in the first place.

[632] By comparison, the US regulator, the FTC, in a report relevant to *all industries* (2012) endorses explicitly the de-identification standard of 'reasonable linkability'.

[633] Health Insurance Portability and Accountability Act of 1996.

[634] Office for Civil Rights (OCR), 2002. Standards for privacy of individually identifiable health information. Final rule. Federal Register, 67(157), p.53181, §164.514 (a).

[635] For that reason, it is more similar to Model 1's proposed jurisdictional legal safe harbour (rather than Model 2), albeit founded upon a model of the identificatory-approach and one that accepts some residual risk of re-identification of a particular person to which data relates in the hands of a recipient.

[636] The following description of that role/task is given: "*A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable: (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and (ii) Documents the methods and results of the analysis that justify such determination".* Of note, the following comments have been made by the OCR in relation to such method in its Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (p.10): "t*here is no explicit numerical level of identification risk that is deemed to universally meet the "very small" level indicated by the method. The ability of a recipient of information to identify an individual (i.e., subject of the information) is dependent on many factors, which an expert will need to take into account while assessing the risk from a data set. This is because the risk of identification that has been determined for one particular data set in the context of a specific environment may not be appropriate for the same data set in a different environment or a different data set in the same environment… No single universal solution addresses all privacy and identifiability issues. Rather, a combination of technical and policy procedures are*

2) **The Safe Harbour Method for De-Identification (§164.514(b)(2)).** This involves the removal of (18) specified individual identifiers of the individual who is a subject of the information (or of relatives, employers, or household members of the individual), as well as absence of actual knowledge by the covered entity that the remaining information could be used alone or in combination with other information to identify the individual.

6.4.2 Devising conditions for obtaining a legal safe harbour based upon effects-analysis

Regarding the nature of the safe harbour conditions set out in EU competition block exemption regulations, as mentioned these are mainly quantitative (like market share or financial thresholds). Conditions of a similar type are found in exemptions to state aid EU rules. The challenge under Model 2 is to devise conditions to enhance certainty relative to the Effects-based Approach per Chapter 4 (and, ultimately, relative to the identificatory-approach status quo and position under the GDPR). Thus, such conditions must be capable of clear specification and interpretation in practice, otherwise low levels of legal certainty may have a chilling effect upon the free flow of data relating to persons. However, conditions devised to be met as indicative of non-appreciable harm likely to flow from a particular processing activity upon a relevant individual do not necessarily have to be quantitative.

Under HIPAA, and the state aid GBER (see fn.602 and reference to its aid-transparency requirement), the terms of a legal safe harbour can involve more qualitative tasks to be carried out by the organisation in order to satisfy the conditions and gain the benefit of the legal safe harbour under offer. A similar process-oriented approach would fit in with current conceptualisations of data protection as a procedural-focused regime. It would also align with 'procedure-based' conceptualisations of personal data, as preferred by scholars such as Hartzog/Rubinstein (2016), focusing on implementing preconditions and (legal/technical/organisational) processes necessary for mitigating the likelihood of future harm befalling the data subject (e.g. guided by sector-specific standards around the types of procedures/methodology that might be used based on common

---

often applied to the de-identification task. OCR does not require a particular process for an expert to use to reach a determination that the risk of identification is very small. However, the Rule does require that the methods and results of the analysis that justify the determination be documented and made available to OCR upon request. The following information is meant to provide covered entities with a general understanding of the de-identification process applied by an expert. It does not provide sufficient detail in statistical or scientific methods to serve as a substitute for working with an expert in de-identification. … Stakeholder input suggests that the determination of identification risk can be a process that consists of a series of steps. First, the expert will evaluate the extent to which the health information can (or cannot) be identified by the anticipated recipients. Second, the expert often will provide guidance to the covered entity or business associate on which statistical or scientific methods can be applied to the health information to mitigate the anticipated risk. The expert will then execute such methods as deemed acceptable by the covered entity or business associate data managers, i.e., the officials responsible for the design and operations of the covered entity's information systems. Finally, the expert will evaluate the identifiability of the resulting health information to confirm that the risk is no more than very small when disclosed to the anticipated recipients."

processing activities within sectors), to ensure sufficient protections are embedded explicitly upfront into the operation of data processing and managed thereafter.

Notwithstanding, any setting of conditions must still overcome the problems per Chapter 4 in relation to defining harm (including considerations around secondary harm, and subjectively-conceived harm) as this must still inform the outcome of a procedure-based effects-centric approach. Indeed**,** what would it take to likely cause harm to data subjects (i.e. liable to affect their privacy/data protection interests in ways that would not be ignored by a reasonable person in a certain processing situation) *appreciably*, as opposed to non-appreciably, and can one or more conditions ever capture succinctly when the 'appreciability' threshold is (likely to be) met? The extent of this challenge magnifies when considering that the conditions would not only have to capture the magnitude of harm, but also associate this with guidance on the requisite standard of likelihood that the harm will materialise when assessed ex-ante.

A natural starting-point for tackling these issues involves more exploration of the underlying rationale for an effects-based approach. In turn, this would require developing coherent and sound theories of harm under such an approach, as has evolved under competition law learning from the field of economics, on the basis of which a methodology (including factors to be taken into account) for assessing the likelihood of harm occurring from a data processing activity could be built, for application in context-specific scenarios.[637]

Yet, at least currently, data protection/privacy law cannot turn to economists for answers to these questions in the same way that competition (and state aid) law look to them to underpin policy decisions. Indeed, some may argue that such assessments should not be 'straitjacketed' by economic considerations only, as non-economic factors are also at play. Moreover, some may argue that economics is not suited to the task, e.g. as it cannot distinguish between harm that focuses on the impact to one person as compared to harm assessed on a very wide scale ranging from an impact on the person concerned by the processing in question to a general societal impact (e.g. loss of social trust), or take into account expectations of privacy or non-rational behaviour more generally. However, as mentioned, such considerations for rationalising impact assessments are not entirely new issues to data protection law, and its foundational justifications have already been made out with references to exhorting PIAs/DPIAs to be carried out by data controllers under EU law, and related template-methodologies endorsed by regulators cited earlier.

To summarise, the most challenging aspect of implementing Model 2 would be determining the conditions for a new block exemption. However, useful analogies and lessons may be drawn here

[637] See fn.429 above for a broad overview of the key factors that could be configured in such a methodology.

with the US HIPAA law which sets out a safe harbour de-identification standard, the achievement of which can be assured through the adoption of specific safeguards. For example, block exemption conditions might also include seeking legal assurances from recipients regarding their intentions for future processing of the relevant data, e.g. documented in contractual terms prohibiting recipients from doing certain things in the future. Still for consideration are what other conditions beyond imposing legal requirements might be suitable for falling within a new legal safe harbour, how to achieve sufficient clarity in describing them in a text, and how their implementation could be demonstrated (if so required) to DPAs and relevant others.[638]

Of course, this all raises the obvious criticism that it may be as onerous to escape data protection obligations as it is to follow them! However, perhaps drawing this conclusion is inevitable to ensure that relevant individuals are protected from non-negligible harm that might be caused to them by the processing of data about them (bearing in mind that the processing of any data relating to them can be assumed, at the very least, to raise the prospect that it may be used in ways that might harm them).

Such questions for further research exploration might be best-suited for consideration alongside work on developing theories of harm and procedures related to the carrying-out of DPIAs, already underway.[639] For example, the WP has already set out tasks for DPAs in relation to the development of policy around risk-based analyses under the GDPR, including "*updating the list of processing which can be considered to present specific risks by essence*", *"developing guidelines on impact assessments and on other accountability tools",* and *"carrying out enforcement action…which may imply challenging risk analysis, impact assessments as well as any other measures carried out by data controllers…"*[640]

[638] As mentioned previously, there are a host of mitigation measures and safeguards which organisations can implement to protect individuals and minimise risks of harm to them. The crucial issue here is selecting the right mitigation measures, which may differ taking into account the context in question requiring choice of appropriate mitigation measures on a case-by-case basis. As a minimum, the conditions could specify that organisations consider the risks of harm involved, the effectiveness of possible measures to mitigate such harms, and the cost of implementing the measures, as well as the reasonable expectations of the relevant individuals.

[639] See, e.g., the ICO (2014), which includes a section entitled, 'What do we mean by privacy?' discussing privacy harm and risk, and a chapter on 'Identifying privacy and related risks'. In other words, advice has already been provided by regulators on PIAs, which provide a foundation upon which to build as directly relevant to guidance on DPIAs. This includes work done by the ICO and CNIL (the French data protection authority) on privacy risk-management methodology. Also important is WP248, a legal summary of which by this author can be found in Knight (2017, New EU guidelines on data protection impact assessments, [online]).

[640] WP 218, 2014: "*13/ Under the forthcoming regulation, the DPAs' role with respect to the risk-based approach will namely consist of: updating the list of processing which can be considered to present specific risks by essence (Article 33 of the draft regulation), developing guidelines on impact assessments and on other accountability tools (as the CNIL and the ICO did with their privacy risk management methodology), carrying out enforcement procedures in case of noncompliance of controllers, which may imply challenging risk analysis, impact assessments as well as any other measures carried out by data controllers, targeting compliance action and enforcement activity on areas of greatest risk".*

While no 'silver bullet' answers should be expected to this challenge, developing an authoritative methodology - to help data controllers determine when appreciable harm is liable to flow from their data processing activities and how to mitigate this risk of harm to a non-appreciable level - may be more important than answering all these questions definitively. Such a methodology would not be of a checklist-prescriptive type for mechanical application, but rather would set out a framework of factors to be analysed (with accompanying explanations of relevance) for undertaking an effects-based assessment process and making decisions based on the results of this process.

6.4.3 Reconciling ex-ante risk-based assessments with the provision of sufficient legal certainty

To some extent, the same inherent underlying challenge of legal certainty exists as described in Chapters 4-5. While all data relating to persons would be deemed personal data under Model 2, to determine whether they fall within the block exemption, data controllers would still have to base their conduct on the likelihood that they satisfy its conditions (taking into account consideration of the specific circumstances at issue) and document this belief. Moreover, this would be an on-going process of assessment depending on the changing data-environment (new risks of harm may arise with time as factors change) and as plans for data processing activities evolve. Said otherwise, data controllers would still be encumbered with carrying-out context-dependent risk-assessments on the basis of the facts existing at any point in time (meaning that assessments would be sensitive to material changes in the facts, such that the exemption rule applies only as long as the conditions are fulfilled and cease to apply when that is no longer the case).[641]

Despite this acknowledgement, this problem remains under any fact-sensitive (dynamic) legal framework of analysis – including identificatory-approach models - recognising the importance of context in determining the value attached to any piece of data for an individual's privacy protection. Indeed, in general, data protection rules predominantly take the form of general principles that require risk-assessments to be carried out on a case-by-case basis by data controllers as part of any organisation's overall legal-compliance strategy (and, ultimately, by the regulators/courts if they need to step in).[642] This fact necessitates that organisations reach sensible and defensible judgements, based on the circumstances of the case in hand, about whether to proceed with data

[641] Booth et al (2004, p.115): "*[w]hether, in actual fact, a particular piece of information is capable of affecting an individual's privacy is determined by the contingent circumstances of the instant case: it is necessary to take account of the specific context. This ideal type expressly recognises that the 'meaning' and 'value' attached to any piece of data for an individual's privacy will be determined by context".*

[642] The WP has already acknowledged this in WP218. More generally, this acknowledgement could also be deemed true of all legal principles which are applicable in many different situations. As Zwenne (2013) points out "*[i]t couldn't be any other way*".

processing in light of any residual risks. Notwithstanding, arguably, laws requiring upfront risk-assessments (and the approaches they advocate, such as the two discussed in this thesis) can still be *more or less effective* depending on whether they provide a consistent decision-making framework with a sufficient degree of accompanying clarity for those who may be subject to them, to eliminate unnecessary ambiguity and confusion.

As mentioned, under Models 1 and 2, this context-dependent defect could be mitigated by providing EU-level guidance – updated on a periodic basis - on how a new block exemption would work and when its conditions would be likely met. Including examples from relatively-stable and typical data processing contexts and discussion of key factors to be considered - such as the type of data, the type of processing activity, and the purpose behind the processing activity – would form constituent parts of the decision-making framework (i.e. the methodology) that could be developed.[643] Hence, regulators would continue to play an essential role in helping data controllers interpret the law and apply it practically supported by objective standards (i.e. clear, predetermined, and objective rules). Compare, for example, 'black lists' 'and 'white lists' issued by MS DPAs (e.g. in Belgium and Germany) setting out specific examples of types of processing operations requiring a DPIA and those that do not. Such publications have more persuasive force than informal guidance (so-called 'comfort letters' or 'letters of negative assurance') provided directly by regulators to organisations, which is the reason, in fact, why competition law block exemptions were introduced in the first place.

6.4.4 Determining the carve-out scope under a modulated approach to data protection obligations

Alongside challenges in overcoming legal uncertainty regarding when a new Model 2 block exemption might apply, there is also the pressing issue of determining which data protection provisions a data controller would be exempt from if they obtain its safe harbour benefits (at least in the sense that they could not be fined by a data protection authority for not following these obligations upon evidence of best endeavours). Moreover, how should the specific carve-outs chosen be justified, in alignment with the general proposition that data processing activity with a

[643] Booth et al (2004, p.16): *"[i]n order to prospectively assess if privacy is likely to be affected, it may be necessary to have regard to the "likely contexts" in order to assess the likely effect that a particular piece of data will have on that privacy. A classificatory model based around effect may then need to assume relatively stable contexts: taking note of what data is usually available to others within a particular context and their possible use in that context, and the various uses to which the data could be put and the impact of such uses upon the individual's privacy. While recognising that such predictions would be inherently fallible. That is not to say however that some kind of predictive judgment could not be made. It is just to emphasise the fallibility of such a judgement".*

non-appreciable risk of harm to the data subject should be subject to a light data protection regime of accountability obligations?[644]

6.4.4.1 Review of EU data protection obligations

Per Chapter 2, controllers bear primary responsibility for ensuring that processing activities are compliant with EU data protection law (see Article 6(2) DPD). Under the GDPR, this principle is mirrored in Article 24 (and Recital 74): the controller is responsible for implementing appropriate technical and organisational measures to ensure and to demonstrate that its processing activities are compliant with the GDPR's requirements. These measures may include implementing an appropriate privacy policy and adherence to approved codes of conduct providing evidence of compliance. Moreover, data controllers are legally obliged to give effect to the rights of data subjects under EU data protection law.[645]

One of the most important requirements that controllers have under data protection law – alongside the fundamental principles per Chapter 2 (legitimacy, data minimisation, purpose limitation, transparency, data integrity, data accuracy) - relates to data security. Controllers should implement appropriate technical and organisational security measures to protect personal data against: accidental or unlawful destruction or accidental loss; alteration; unauthorised disclosure or access (Article 17(1) and Recital 46 DPD, mirrored in Article 32 and Recital 83 GDPR). Whereas the DPD leaves much discretion to controllers in terms of the measures they can implement suitable

[644] Compare the statement by the WP in Annex to Letter to Mr Paul TIMMERS Director of Sustainable and Secure Society Directorate DG Connect, Brussels, 25 February 2015: "*[t]he Working Party recalls its Statement on the role of a risk-based approach in data protection legal frameworks of 30 May 2014, where it has underlined that fundamental principles applicable to the controllers (i.e. legitimacy, data minimization, purpose limitation, transparency, data integrity, data accuracy) should remain the same, whatever the processing and the risks for the data subjects. The Working Party simultaneously expressed its concern about the introduction of the notion of a lighter data protection regime for pseudonymised data. While pseudonymisation can represent an important safeguard with regard to for example data security, the use of pseudonymous or pseudonymised data is, in itself, not sufficient to justify a lighter regime on accountability obligations".* However, a lighter regime on accountability obligations might be compared with existing rule derogations available to data controllers where personal data are being processed for historical, statistical or scientific ('research') purposes. For example, Article 6(1)(b) DPD contains a specific provision on further processing for these purposes providing a partial exemption from the purpose limitation principle. Such processing shall not be considered as incompatible with the initial purposes for which the personal data was processed provided that Member States "*provide appropriate safeguards*". Recital 29 DPD further provides that "*these safeguards must in particular rule out the use of the data in support of measures or decisions regarding any particular individual*", which suggests that the objective behind requiring such safeguards is to minimise risk of harm arising to the data subjects. Another example relates to Article 11 DPD ('Information where the data have not been obtained from the data subject'), which provides an exemption from the requirement to supply certain notice to the data subject in the case of processing for research purposes where "t*he provision of such information proves impossible or would involve a disproportionate effort or if recording or disclosure is expressly laid down by law. In these cases Member States shall provide appropriate safeguards".* The GDPR contains similarly-worded derogations.

[645] This fact is implied under the DPD, whereas the GDPR formalises this de facto position, see Article 12(2) and Recital 59.

to the data-context, the GDPR is more prescriptive.[646] However, under both regimes, the primary requirement is that the controller must ensure the security of the personal data that it processes.

Also per Chapter 2, EU data protection law creates a host of data subject rights that data controllers must uphold when they process personal data. A key GDPR objective is to strengthen such rights. In terms of the legal obligations that data controllers must undertake to give effect to them, they include the following:

- **Provision of information** - To ensure the fair processing of personal data, EU data protection law obliges controllers to provide minimum sets of information to data subjects regarding the processing of their personal data (Articles 10-11, Recital 38 DPD). Each set should include information on the controller's identity, their reasons for processing personal data, and other relevant information necessary to ensure the fair processing of personal data upon first collection and further re-use. Under Articles 5(1), 12-14 GDPR (see also its Recitals 39, 58 and 60), controllers must supply similar minimum information to data subjects using plain and clear language, in a format that is concise, transparent, intelligible and easily accessible.

- **Right of access** – Per chapter 2, controllers are also obliged to enable data subjects to enforce subject access rights with respect to their personal data (albeit there are some exceptions and restrictions set out in Article 13, DPD). Broadly, the reason behind giving subject access right is to enable an individual to check whether their data is being processed lawfully and to verify the accuracy of such data.[647] The GDPR expands the mandatory categories of information already required to be supplied under the DPD in connection with a data subject access request.[648]

---

[646] Depending on the nature of the processing, these measures may include: encryption of the personal data; on-going reviews of security measures; redundancy and back-up facilities; and, regular security testing. Adherence to an approved code of conduct may provide evidence that the controller has met these obligations.

[647] Case C-553/07, College van burgemeester en wethouders van Rotterdam. v. M.E.E. Rijkeboer [2009] ECR I3889 (para.49): "*[t]hat right to privacy means that the data subject may be certain that his personal data are processed in a correct and lawful manner, that is to say, in particular, that the basic data regarding him are accurate and that they are disclosed to authorised recipients. In other words, rights are given to data subjects because they have an interest in the protection of their fundamental rights and freedoms, especially their right to privacy*". To this end, the CJEU in Joined Cases C-141/12 and C-372/12, YS v Minister voor Immigratie**,** Integratie en Asiel [2014] EU:C:2014:2081 examining the 'relating to' element of the personal data concept – justified its conclusion that a piece of legal analysis was not personal data because in the factual context it would not "*serve the Directive's purpose*" to find it so (para. 46): "*extending the right of access of the applicant for a residence permit to that legal analysis would not in fact serve the Directive's purpose of guaranteeing the protection of the applicant's right to privacy with regard to the processing of data relating to him, but would serve the purpose of guaranteeing him a right of access to administrative documents, which is not however covered by Directive 95/46".*

[648] Under the DPD (Article 12(a), Recitals 27, 41-44), data subjects have the right to obtain: confirmation of whether the controller is processing their personal data; information about the purposes of the processing; information about the categories of data being processed; information about the categories/identities of recipients with whom the data may be shared; a copy of those data (in an intelligible format) and information as to the source of those data; and an explanation of the logic involved in any automated processing that has a significant effect on data subjects. Under the

- **Right of rectification –** Under the DPD (Articles 6(1)(d) and 12(b)), controllers should ensure that inaccurate or incomplete data are erased or rectified. Data subjects have the right to rectification of their personal data where the controller fails to comply with the DPD. The position under the GDPR (Articles 5(1)(d) and 16, Recitals 39, 59, 65, and 73) is largely equivalent in respect of controller obligations - they should ensure that inaccurate or incomplete data are erased or rectified, and data subjects should be entitled to demand the rectification of inaccurate personal data.

- **Right to erasure -** Under the DPD (Article 12(b)), data subjects can request that their personal data be deleted where the controller fails to comply with its data protection requirements and the continued processing of such data would be unjustified. The GDPR creates a broader right to erasure (previously entitled, a 'right to be forgotten'). Under its Article 17 (see also Recitals 65-66, and 68), data subjects have such right to erasure of personal data if: the data are no longer needed for their original purpose (and no new lawful purpose exists); the lawful basis for the processing is the data subject's consent, the data subject withdraws that consent, *and* no other legal basis exists; the data subject exercises the right to object, and the controller has no overriding grounds for continuing the processing; the data was processed unlawfully; or erasure is necessary to comply with EU law or MS laws.[649]

- **Right to restrict processing –** Under the GDPR, data subjects can require that personal data processing be limited to certain purposes (e.g., where the personal data is no longer needed by the controller for its original purpose, but still required by them to establish, exercise or defend legal claims).

- **Notifying third parties –** Controllers must inform any third parties with whom they have shared personal data when its data subjects have exercised rights of rectification, erasure, or blocking (Article 12(c) DPD), unless such notification would be impossible or require disproportionate effort. The GDPR contains similar obligations (in Articles 17(2) and 19,

---

GDPR (Article 14, Recital 63), a data subject has the right to obtain: confirmation of whether, and – if so - where, the controller is processing their personal data; information about processing purposes; information about the data categories being processed; information about the categories of recipients with whom their personal data may be shared; information about the period for which the data will be stored (or the criteria used to determine that period); information about the existence of the right to complain to a data protection authority; information about the existence of a right to erasure, a right to rectification, a right to restrict processing, and a right to object to processing; where the data was not collected from the data subject, information as to its source; and, explanation of the logic involved in, any automated processing that has a significant effect on the data subject. Additionally, data subjects may request a copy of the personal data being processed. Moreover, under the GDPR (Articles 12(5), 15(3)-(4), Recital 59), the controller must give effect to the right of access free of charge (with limited exceptions).

[649] The GDPR also exempts research from the right to erasure insofar as it is "*likely to render impossible or seriously impair the achievement of the [research] objectives*" (Article 17(3)(d)).

Recital 62), which require controllers to implement procedures/systems for notifying affected third parties about the exercise of those rights.

- **Right of data portability –** Whereas the DPD does not directly acknowledge a right of data portability, the GDPR (Article 20, Recitals 68 and 73) gives data subjects the right to transfer personal data relating to them between controllers. In that context, data subjects have a right to receive a copy of such data in a commonly-used, machine-readable format that can be transferred directly between controllers.

- **Right to object to processing –** Per Chapter 2, a controller must have a legal basis to carry out each act of personal data processing. Where that legal basis is either 'public interest' or 'legitimate interests of the controller' (see fn.51 above), data subjects have a right to object based on any compelling legitimate grounds to such processing (Article 14(a) and Recitals 30 and 45, DPD) and – if their objection is found to be well-founded - the controller must stop the relevant processing activity in relation to those data. Under the GDPR (Article 21(1)),[650] by comparison, if a data subject exercises his/her right to object, controllers bear the upfront onus of showing they either have compelling grounds for continuing processing which overrides the interests, rights and freedoms of the data subject, or that the processing is necessary in connection with establishing/exercising/defending his/her legal rights. Otherwise, they must stop the processing activity.[651]

- **Right to not be evaluated on the basis of automated processing –** Under the DPD (Article 15, Recitals 11, 15, 27 and 41), data subjects are entitled not to be subjected to decisions based solely on automated processing of personal data for the purposes of personal evaluation. Exceptions to this rule are: where the processing is performed in the course of entering into contracts with data subjects, provided that appropriate safeguards are in place; or, where the processing is authorised by law. The GDPR preserves this position with minor amendments (Article 22, Recitals 71 and 75). As previously mentioned, controllers data subjects have a right not to be subject to (generally construed as a prohibition on taking) decisions based solely on automated processing with significant/legal effects. Such

[650] Article 21(1) GDPR: "*[t]he data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims".*

[651] Other GDPR obligations related to the right to object include: an obligation to inform data subjects of the right to object (Articles 13(2)(b), 14(2)(c), 15(1)(e), 21(4)); a right to object to processing for the purposes of direct marketing (Article 21(2)-(3), Recital 30); and, a right to object to processing for scientific, historical or statistical purposes unless the processing is necessary for the performance of a task carried out for reasons of public interest (Articles 21(6) and 83(1), and Recital 156).

processing is only permitted in three scenarios: where it is necessary for entering into or performing a contract with the data subject provided that appropriate safeguards are in place; where it is authorised by law; or, where the data subject has consented explicitly to it and appropriate safeguards are put in place.[652]

6.4.4.2 Exemptions under EU data protection law regarding the carrying-out of certain data controller obligations as comparative exemplars

There are examples of exemptions/derogations from data protection obligations in the DPD and the GDPR. Under the DPD, for example, exemptions apply in respect of processing situations where personal data are used solely for journalistic purpose, or for artistic or literary expression, in balancing privacy with freedom of expression. The exempted obligations include information provision and access duties. However, exemptions must be applied in accordance with law and must respect the principle of proportionality in a democratic society.[653]

As mentioned in Chapter 5, the GDPR's exemptions from controller obligations include those available to those who process personal data for scientific and historical research purposes (compare fn.644 above regarding the existing DPD research exemption): to avoid restrictions on storage limitation; on processing sensitive categories of data; and on secondary processing so as to allow them to further process personal data beyond the purposes for which they were first collected (Article 6(4), Recital 50).[654] To benefit from these exemptions, researchers must

---

[652] Article 22(1) GDPR (compare Recital 52) also prohibits controllers from subjecting a data subject to a decision "*based solely on automated processing, including profiling*" as a result of processing personal sensitive data (as defined in Article 9 of the GDPR), except in limited circumstances. There must be "suitable safeguards" in place and the processing must be based on Article 9(2)(a) or (g), requiring the data subject's explicit consent or substantial public interest (Article 22(4)). If the primary purpose for initially collecting sensitive data was for research under Article 9(2)(j), or if the data was "*manifestly made public by the data subject*" under Article 9(2)(e), profiling is prohibited. The exception, under Article 9(2)(j), allows a researcher to process sensitive data where, "*processing is necessary for [research] purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject*."

[653] Domestic laws can also provide exemptions from some of the DPD's provisions in matters of: national security and defence; the prevention, investigation, detection and prosecution of criminal offences; the protection of data subjects and the rights and freedom of others.

[654] When a controller collects personal data under a lawful basis, such as consent, Article 6(4) GDPR allows it to process the data for a secondary research purpose because it allows for subsequent processing operations that are compatible and Recital 50 specifies that further processing for research purposes "*should be considered to be compatible*". Article 5(1)(b) also states, "*further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes*." Article 89 GDPR (quoted in the footnote directly below) sets out the safeguards that controllers must implement in order to further process personal data for research. The GDPR also creates additional safeguards to protect individuals from certain intrusive types of processing. Article 35(2)(a) requires controllers to conduct a PIA any time "*a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person*." By contrast, under the DPD, the presumption is that a controller cannot further process personal data beyond the purposes for which it was originally collected, unless the relevant MSs have enacted legislation permitting such processing activities for research purposes (such as under section 33 of the DPA).

implement appropriate safeguards (including potentially pseudonymisation),[655] in keeping with recognised ethical standards, that lower the risks to the rights of individuals. As long as they implement such safeguards, moreover, the GDPR also carves-out exemptions from certain data subject rights in this processing context.[656]

6.4.4.3 Which data protection obligation exemptions might be appropriate under Model 2?

One possible starting point for evaluating this issue is to run through all of the controller's obligations to eliminate *inapplicabl*e categories of potentially exempt obligations (i.e. because the *non-realisation* of them in relation to a particular processing activity would always imply the prospect of appreciable harm to the data subject concerned).

One obvious candidate for retention is security requirements.[657] If personal data is not subject to adequate security at one time stage (upon one processing activity), this raises sharply the likely risk of secondary harm befalling the data subject at a later stage (including upon additional processing activities) resulting from compromised data relating to persons in perpetuity. Specifically this harm may originate from third parties' actions flowing from inadequate data security being in place.[658]

---

[655] Article 89 GDPR (regarding safeguards and derogations for the processing of personal data for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes) states: "*[p]rocessing of personal data for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes, shall be subject to in accordance with this Regulation appropriate safeguards for the rights and freedoms of the data subject. These safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure the respect of the principle of data minimisation. These measures may include pseudonymisation, as long as these purposes can be fulfilled in this manner. Whenever these purposes can be fulfilled by further processing of data which does not permit or not any longer permit the identification of data subjects these purposes shall be fulfilled in this manner".*

[656] Under Article 21 GDPR, data subjects retain a right to object to processing, even for research purposes. However, (per fn.651 above), a researcher may override a data subject's objection if "*the processing is necessary for the performance of a task carried out for reasons of public interest*" (Article 21(6)). As mentioned (per fn.649 above), the GDPR also exempts research from the right to erasure insofar as it is "*likely to render impossible or seriously impair the achievement of the [research] objectives*" (Article 17(3)(d)). Additionally, EU MSs may craft exemptions to a number of other rights by appropriate legislation in relation to personal data processed for research purposes (Article 89 GDPR): *"[w]here personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21* ***subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.*** *Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21* ***subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.*** *Where processing referred to in paragraphs 2 and 3 serves at the same time another purpose, the derogations shall apply only to processing for the purposes referred to in those paragraphs)"* (emphasis added).

[657] ICO (2013, pp. 6-7): *"[w]e prefer a wide definition of personal data, including pseudonymised data, provided the rules of data protection are applied realistically,* ***for example security requirements*** *but not subject access….However, it is important to be clear that a wide definition plus all the associated rules in full would not work in practice. This is a real issue in contexts as diverse as medical research and online content delivery"* (emphasis added).

[658] See, e.g., the WP's comments (in WP217, p.38): "*access to the Internet, exchanges of data with sites outside the EU, interconnections with other systems and a high degree of system heterogeneity or variability can represent vulnerabilities that hackers could exploit. This risk source bears a relatively high likelihood for the risk of compromising data to materialise. Conversely, a homogeneous, stable system that has no interconnections and is disconnected from*

An alternative starting point is to run through all of the controller's obligations under the DPD/GDPR to consider one-by-one which one's non-application (exemption) *may* be justifiable, in situations where their *application* would be disproportionate to a level of non-appreciable harm-risk to data subjects from data use. For example, the justification may be that that application of particular obligations would perhaps decrease, rather than increase, the possibility of privacy harm befalling the individual (such as by requiring that more information be associated with them).[659]

An inroad into this starting point discussion is reconsideration of Article 11 GDPR ('processing which does not require identification').[660] As mentioned, data controllers have a legal obligation to give effect to the rights of data subjects under EU data protection law.[661] An exception is provided by Article 11 in respect of the rights described in Articles 15-20 GDPR,[662] in situations where the controller is not in a position to identify the data subject from data in its possession (at least to the extent that it can *demonstrate* that it is not in a position to identify the data subject).[663] In that same context, the controller is also not obliged under Article 11 to seek out further information in order to link data in its possession to a data subject solely for compliance with data protection rules. Otherwise, it would create the perverse result of obligating controllers to collect more personal data just in order to identify data subjects for the mere purpose of complying with the GDPR.

This logic behind Article 11's inclusion makes sense on the assumption that, if someone wants to make use of his/her subject access rights, the controller has to be able to identify the one requesting access. Yet the possible reasons *underlying* this argument are also worth examining, including:

---

the Internet bears a far lower likelihood of compromising data". Of course, this may not necessarily equate to risk of harm depending on the sensitivity of the data and the further uses to which it may be put (although such uses are effectively unpredictable because a third-party data hacker would be involved, and so can be assumed to involve likely appreciable harm to the data subject).

[659] Another justification used by Solove/Schwartz for a modulated approach is that the application of certain data controller obligations can cripple socially productive uses of analytics in ways unjustified where they do not raise significant risks of individual privacy harms. See, Schwartz & Solove (2011, p. 1880): *"[m]oreover, limits on information use, data minimization, and restrictions on information disclosure should not be applied across the board to identifiable information. Such limits would be disproportionate to risks from data use and would cripple socially productive uses of analytics that do not raise significant risks of individual privacy harms. Some of these uses of analytics are consumer-oriented and some are not, but the benefit to the public is often clear".*

[660] GDPR, Article 11: "***Processing which does not require identification*** *1.If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation. 2. Where, in cases referred to in paragraph 1 of this Article, the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 20 shall not apply except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification".*

[661] Article 12(2) and Recital 59 GDPR.

[662] These are respectively, the right of access by the data subject (15), the right to rectification (16), the right to erasure (17), the right to restriction of processing (18), notification obligation regarding rectification or erasure of personal data or restriction of processing (19), right to data portability (20).

[663] In other words, Article 11 implies that, from the moment a data subject can be identified, the exercise of their rights should be made possible. To note, the GDPR permits controllers to require proof of data subjects' identities before giving effect to their rights and controllers must use all reasonable efforts to verify these identifies upfront (see Article 12(2) and (6) and Recital 57 GDPR),.

1) to let data controllers know who is trying to exercise their rights (and indirectly, therefore, to avoid third parties defrauding the controller by impersonating the data subject),[664] and/or

2) to avoid requiring that data controllers pro-actively (re-)identify data subjects by requesting more information than they already have in their possession (and thereby diminishing their privacy).

Yet, as Solove/Schwartz point out in commenting on Article 11 in draft form, this Article's provisions are in fact "vague" in terms of when it applies and the consequences of its application.[665] In respect of reason (1), part of the vagueness may be attributed to confusion over interpreting the term 'to identify' in this context (to revisit the arguments per Chapters 1 and 3). Reason (1) suggests that it need only be equated with being able to authenticate the data subject (confirming that the person is the same person to whom a particular piece of information relates), rather than knowing their 'real-world' identity.[666] Said otherwise, there is an argument that – where a controller 'identifies' a data subject by singling them out such as on the basis of digital identifiers – data subjects should be entitled to authenticate themselves via that medium (without necessarily disclosing their real-word identity in full), as long as controllers can enable data subjects rights to be fulfilled on that basis. Therefore, we may be able to discount reason (1) as a justification for Article 11 GDPR to some extent.

Regarding reason (2), in comparison, the possibility for privacy risks is in fact linked to the increased possibility of privacy harm befalling the data subject when a controller stores personal data in an identified (not just an identifiable form).[667] In other words, Article 11 may be seen as *being aimed*

---

[664] Schwartz & Solove (2013, #1, p.915): "*entities would need to maintain an ongoing connection between the individual and the identifiable information to allow that individual to exercise her rights of notice, access, and correction. In this fashion, the law's implementation could force the transformation of identifiable data into identified data. Article 10 of the Proposed Regulation explicitly seeks to avoid this result…Administering certain FIPs] requires that data be identified… Providing individuals with access to their data, for example, requires that the information be kept in identified form…* ***If data is not kept in this form, data processors would not know to whom to provide access****"* (emphasis added).

[665] Ibid, p.913: "*…vague regarding (1) the types of personal data to which it would apply and (2) the provisions of the regulations with which a data controller need not comply if it had such information".*

[666] Although, the opposite argument is implied by Zwenne (2013, p.9): "*[i]f, for example, someone wants to make use of his or her subject access rights, the controller has to establish the identity of the one requesting access. How can that be done when the data subjects are not known? This will be difficult, if not downright impossible - when it concerns access to data about individuals whose identity is unknown. What is the value of access rights in such a situation? And what about informing the data subjects, one of the core obligations under the law?"*

[667] Tene (2011, p.7): "*the nature of data as personal or not could be viewed as a continuum, as opposed to the current dichotomy. This means that data which are only identifiable at great cost would remain within the legal framework, yet be subject to only a subset of fair information principles. Hence, for example, it makes little sense to provide individuals with a right of access and rectification to data that are not readily identifiable, as this would require data controllers to proactively re-identify data, infringing the privacy of individuals requesting access and others".* Compare, Schwartz & Solove (2013, #1, p.915): "*[a]dministering certain FIPs requires that data be identified, and keeping data in identified format can create privacy risks. Providing individuals with access to their data, for example, requires that the information be kept in identified form. But by keeping the data in identified form, there is an increased risk from a*

at incentivising the processing of data relating to persons in ways that reduce the likelihood of privacy harm flowing from that activity to those individuals, a condition of which may be not storing data relating to persons from which they can be directly identified. Perhaps, therefore, the Article 11 GDPR exemption could be better reframed as a benefit that is justified *precisely because* the likelihood of harm occurring to the data subject is not appreciable. Conversely, based on the same logic as Article 11 was introduced, exempting obligations upon data controllers to fulfil data subject rights under Model 2 might similarly be deemed appropriate.

Notwithstanding some suggestions are made in the case scenario discussions above, supported by arguments by Balboni et al, it is considered outside the scope of this thesis to conclude upon what rules might be carved out under a new Model 2 block exemption regulation, apart to stress that this requires very careful consideration. This research is worthy of a separate study and methodological development, not least in light of the need to assess fully counter-objections to a lighter regime of data protection accountability obligations, such as that its introduction might lead to a level of privacy protection that is too low.

However, criticisms by the WP, which argues that all data subjects should have the same level of protection under data protection rules can be rebuffed.[668] For example, the WP has commented that, "*controllers should always be accountable for compliance with data protection obligations including demonstrating compliance regarding any data processing whatever the nature, scope, context, purposes of the processing and the risks for data subjects"* and *"[f]undamental principles applicable to the controllers (i.e. legitimacy, data minimization, purpose limitation, transparency, data integrity, data accuracy) should remain the same, whatever the processing and the risks for the data subjects"*.[669] Yet this argument that the introduction of a modulated data protection

---

potential data security breach. If data is not kept in this form, data processors would not know to whom to provide access… By contrast, when a breach involves only identifiable data, the harm that the information can cause to individuals is much less likely to occur. Although identification of some data may be theoretically possible, individuals with unauthorized access to it may lack the resources or knowledge to do so...keeping data in de-identified form prevents harms from inappropriate access by employees or others… ***Harm can only occur when the party who obtains the data also knows how to identify it"*** (emphasis added).

[668] See Article 29 Working Party (Statement of 27 February 2013, p.3): *"[b]asing exceptions on quantitative qualifiers risks excluding companies from certain obligations that are actually of vital importance. Data subjects should have the same level of protection, regardless of the size of the organisation or the amount of data it processes. Therefore the Working Party feels that all controllers must act in compliance with the law, though this can be done on in a scalable manner." "2/ Rights granted to the data subject by EU law should be respected regardless of the level of the risks which the latter incur through the data processing involved (e.g. right of access, rectification, erasure, objection, transparency, right to be forgotten, right to data portability)".* In this context, the WP is expressing concern that - both in relation to discussions on the GDPR, and more widely - references to 'a risk-based approach' is being increasingly and wrongly presented as an alternative to well-established data protection rights and principles, rather than as a scalable and proportionate approach to compliance.

[669] WP 218, points 3/ and 4/. Albeit the WP also says in that statement, *"[t]here can be different levels of accountability obligations depending on the risk posed by the processing in question*" and *"[t]his means that a data controller whose processing is relatively low risk may not have to do as much to comply with its legal obligations as a data controller whose processing is high-risk".* While the application of the data protection principles are deemed *"inherently scalable*" taking into account personal data processing's nature and scope (including *"the categories of the data it processes"*),

introduction might lead to a too-low level of privacy protection, is countered by the fact that all data relating to a particular living individual would likely be deemed personal data in the future (as the identificatory requirement becomes interpreted widely, including under Model 2). Consequently, privacy protection is not reduced compared with the situation under the GDPR, rather it is liable to be enhanced in those scenario-circumstances where it is deemed most appropriate in light of the potential risks. Furthermore, Model 2 makes practical sense precisely because the alternative – the application of the full gamut of data protection rules to the processing of all data about people in the future - will become more and more unworkable in practice.[670]

Lastly, AG Kokott in her 2017 Opinion on the *Nowak* case raises an interesting point in justification of a modulated approach in practice:

> [T]he classification of information as personal data cannot be dependent on whether there are specific provisions about access to this information which might apply in addition to the right of access or instead of it. Further, neither can problems connected with the right of rectification be decisive in determining whether there exists personal data. If those factors were regarded as determinative, certain personal data could be excluded from the entire protective system of the Data Protection Directive, even though the rules applicable in their place do not ensure equivalent protection but fragmentary protection at best.[671]

In other words, she suggests that the legal classification of personal data is independent from whether – in the circumstances – *all* data subject rights can be enforced in respect of its processing (as long as the most important ones can).[672]

---

the WP also comments, *"[i]mplementation of controllers' obligations through accountability tools and measures (e.g. impact assessment, data protection by design, data breach notification, security measures, certifications) can and should be varied according to the type of processing and the privacy risks for data subjects*"…*There should be recognition that not every accountability obligation is necessary in every case – for example where processing is small-scale, simple* ***and low-risk****. The form of documentation of the processing activities can differ according to the risk posed by the processing… Yet, all data controllers should at least to some extent document their processing activities in order to further transparency and accountability. Documentation is an indispensable internal tool for controllers to manage accountability effectively and for ex-post control by DPAs as well as for the exercise of rights by data subjects. It goes beyond information to be given to the data subjects"* (emphasis added).

[670] ICO (2013, pp. 6-7): "*[w]e prefer a wide definition of personal data, including pseudonymised data, provided the rules of data protection are applied realistically, for example security requirements but not subject access….However, it is important to be clear that a wide definition plus all the associated rules in full would not work in practice. This is a real issue in contexts as diverse as medical research and online content delivery*".

[671] C-434/16, Nowak v. Data Protection Commissioner, Opinion of Advocate General Kokott delivered on 20 July 2017, para 34. For the facts of this case, see fn.124 above.

[672] What is ultimately important, to the AG, in this context, it seems is enabling data subjects to uphold any rights in relation to which they have an legitimate purpose based on the protection of his/her private life (para 26). For example, she states (para 39), "*even irrespective of rectification, erasure or blocking, data subjects generally have a legitimate interest in finding out what information about them is processed by the controller"* and (para 41) *"must at least be able to find out whether his script is still being retained".* Using a teleological approach to the concept of personal data, therefore, the AG suggests that the ability for data subjects to act upon these core interests "*presupposes that the incorporation of the examination candidate's personal data in the script is recognised*". This raises the question, in turn,

6.4.5 Emphasising the possibility for non-interference of Model 2 with the requirement for ensuring legal basis to justify personal data processing and related analysis

Notwithstanding the open-endedness of the above conclusions, in light of possible arguments that jurisdictional issues would risk becoming confused with certain analyses for carrying-out - in complying with substantive data protection law provisions - under Model 2, it should be emphasised that there is no essential incompatibility between it and the current workings of the 'criteria for lawful processing of personal data' (legal bases) regime under Article 7 DPD (and Article 6 GDPR) per Chapter 2.[673]

Despite the recommendations in the case scenario in section 6.3 above, it is not necessarily intended that Model 2 would exempt data controllers from either of the following requirements:

- to obtain consent from data subjects for particular processing activities to be carried out on personal data relating to them, or alternatively, to ensure that another legal basis for carrying-out personal data processing to be relied upon in the alternate is available (e.g. legitimate interest); or

- to ensure compatibility of purpose between an original and a further processing act assuming that the same data, if deemed on both occasions to be personal data, are to be repurposed.[674]

However, as the case study discussion above recognises, Model 2 could provide a helpful departure point from which to avoid some of the problems arising from demonstrating a new lawful basis for secondary processing in certain cases where it is not practical. Compare the proposal put forward by the European Parliament Rapporteur (Jan Albrecht) during the GDPR reform discussions that, where personal data are processed in the form of pseudonyms, there should be an exemption from

---

whether these core interests – i.e. the right to find out what personal data about is being processed by a controller and whether it is still retained - be the notional 'yardstick' against which a decision could be taken as to which controller obligations might be excluded under a new Model 2 block exemption regulation, with no possibility for exemption of at least these core data subject rights. However, the problem with this proposition is that, if the answer is 'yes' and such information is currently being stored about you, to exempt the remainder rights at your disposal to implement changes in relation to such rights would have to be strongly justifiable, at least with regards to the context of the processing. This problem consideration would require future research.

[673] As a reminder (see fn.51 above), this Article sets out the legal (legitimising) bases at least one of which must be satisfied for personal data processing to be lawful. In other words, it focuses on the criteria that must be satisfied in order for personal data processing to be justified in the first place (i.e. the legal grounds for justification).

[674] Not least this is because, even when data processing is deemed lawful, it can still have implications for the privacy of the data subject. Thus, while the legitimising ground relied upon by the data controller may be a relevant factor in considering the likely effects of a processing activity upon a data subject (such as to help assess reasonable data expectations, or what the intention of the data controller is), it is not the only one and it is liable to change over time. Although, to note, there is an unresolved conceptual issue not so far mentioned about whether the further processing of data relating to persons (where it was *not* previously deemed personal data in relation to an earlier processing activity) would still have to be compatible with that earlier processing purpose. This issue is left open for future research.

obtaining data subject consent except by automated means using an EU-valid technical standard (such as 'do not track') (Article 7, paragraph 2A (new)). Compare also the approach proposed by Gratton (2013, p.114f) that information qualifying as personal would be managed in light of its overall sensitivity, such that less stringent consent would be required before collecting/using/disclosing data that presents a lower risk of harm. Ultimately, regulators could inform that decision.

Finally, Model 2 is not intended to overlap with Article 7 DPD and its subsection (f) (mirrored in Schedule 2, section 6(1) DPA), updated into Article 6(1)(f) of the GDPR.[675] Article 7(f) is the 'legitimate interest' legal basis upon which data processing may take place and involves a balancing test. It requires that the rights of the data subject are balanced against the legitimate interests of the data controller. Its linguistic formula might encourage organisations to self-assess likely processing effects in a strict cost-benefit analysis way under Model 2, in turn giving rise to confusion between, and conflation of, tests. To repeat earlier discussion, however, Model 2 does not involve an explicit balancing exercise to be carried out (i.e. balancing the rights of the data subject in terms of harm that might result to them from a data processing activity, and the data controller's interests in carrying out such an activity – something that can end up in a very subjective determination). Model 2 does not essentially relegate the law to an exercise whereby data controllers weigh up and trade-off pros and cons, but it does get closer to figuring out how to shift focus onto recognising socially beneficial uses of big data without glossing over the harms that can simultaneously result in a word where more data are inevitably processed about people (directly or indirectly).

Moreover, while Model 2 is conceived to act independently of the legitimate interest test, per the suggestion in the case study discussion there may be merit in considering an exemption to the requirement of satisfying the legitimate interest test in certain cases (where no other lawful basis is relevant).

6.5 Chapter conclusion

This chapter developed the effects-based exemption proposition introduced in Chapter 5, with the aim of finding ways in which the use of an exemptions model might improve the data protection regulatory regime in practice focusing on two areas. These are: to increase legal certainty associated with the practical application of the approach (not least by improving coherency with

[675] It states: "*Member States shall provide that personal data may be processed only if… (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection under Article 1(1)*".

other facets of the data protection regulatory regime that currently work well); and, to strengthen regulatory incentives to ensure compliance in practice with carrying-out effects-based analyses – and encourage the mitigation of harm - as part of jurisdictional considerations under EU data protection law.[676] In turn, this should help develop a more coherent theoretical effects-based model in response to the challenges described.

In such a manner, this Chapter has addressed sub-research question 3:

> **Can the practical disadvantages associated with an effects-based approach to personal data be ameliorated by the use of block exemption provisions, as exemplified under EU competition law (a distinct area of law and regulatory system that has been modernised using an effects-based policy approach)?**

In concluding how the discussion has answered that question, the following summary key points can be drawn:

- Analogies may be drawn between the VBER and its accompanying Vertical Restraints Guidelines under EU competition law and proposals made in this chapter for improving legal certainty under an effects-based approach under EU data protection law. They both necessarily depend on a firm foundation of policy guidance addressed at decision-takers/makers in practice. They encompass an assessment-methodology setting out a framework of analysis to be followed and factors for consideration, with practical examples on how organisations can satisfy - and demonstrate satisfaction - of the terms of a new legal safe harbour in the eyes of regulators. Additionally, discussion of remedial actions that can be taken in order to minimise negative effects flowing from their processing activities is valuable. Thus, guidance accompanying a potential new data protection block exemption regulation should have regard to relatively typical processing contexts that may arise – particularly, clearly beneficial ones - and, more widely, the factors to be considered to prospectively assess the effects that a particular type of processing activity upon a particular type of data might have on a data subject or other collective interests.[677]

---

[676] In fact, as mentioned, potential success in improving both factors seems inter-related to some extent. Whereas, the more legal certainty that a data controller has in knowing how the law might apply to their data processing activities, the more incentive they may have in carrying them out (and vice versa).

[677] Booth et al (2004, p.16): "*[a] classificatory model based around effect may then need to assume relatively stable contexts: taking note of what data is usually available to others within a particular context and their possible use in that context, and the various uses to which the data could be put and the impact of such uses upon the individual's privacy… It must still be emphasised however that, while one may anticipate a particular data type as more or less likely to affect an individual's privacy in a particular way, this prediction is inherently fallible.* ***If one wishes to assess the relative impact that various data types are likely to have upon an individual's privacy then it may be necessary to assume a stable context. That context will have to take note of not only what data is usually available to others within that particular context, but also how those various pieces of data could ordinarily be used within that context. The***

- Notwithstanding, it cannot be side-stepped that there are challenges with respect to providing clearly-scoped, effects-based conditions requiring satisfaction so that the safe harbour benefits of a new block exemption regulation might be availed upon by those processing data relating to persons.[678] Similarly, the condition-types found in EU competition/state aid law block exemptions are not easily copied. Not least, this relates to the fact that – as yet - there are no clear economic (or other scientifically-measured) theories of harm when it comes to data protection law. For example, competition law talks about consumers (and consumer harm) as a whole, rather than individual consumers affected by contingent circumstances and subjective considerations in any instant case.

- Improving regulatory incentives is also at the heart of safe harbour provisions – including those discussed in this chapter. Under both Models 1 and 2, for example, the regulatory driver would be to incentivise organisations to reduce the likelihood of data subjects being harmed by their intended data processing activities in advance of executing such plans. Whereas the prospect of obtaining compliance obligation exemptions would be the commercial driver for organisations taking actions to satisfy block exemption conditions. Both strands are consistent, essentially, with key features of existing and future EU data protection law. For example, the principle of organisational self-certification associated with safe harbour provisions blends well with self-assessment requirements under the DPD. Moreover, a central plank of the GDPR is its promotion of the principle of organisational accountability – where the organisation processing the data bears the responsibility to demonstrate compliance with data protection law obligations - through the adoption of certain technical and organisational measures and tools (e.g. DPIAs, 'data protection by design and default', data breach notification, security measures, and certifications) and related documentation. In other words, it places the burden of data protection squarely upon those who process data relating to persons (that may well be deemed personal data, if not now but at a later date) to minimise negative consequences. Indeed, these may be seen, rightly, as part of the same process of transparent risk-assessment - of the type exhorted by policy makers as part of PIAs and DPIAs analysing likely effects of different

informational context will have to take account of the relationship between the data in question and the participants (both data subject and data controller). A classificatory model constructed around 'affect' will then have to take note of the various uses to which different data types may be put (by all involved) and the impact that such uses may have upon an individual's privacy" (emphasis added).

[678] To note, in competition law, the Commission's approach is to issue a block exemption only after it has gained experience of a particular category of agreement through past policy-decision making and knowledge gathering. This then enables the Commission to identify the circumstances in which the conditions for exemption in Article 101(3) are likely to be met and to define those circumstances in the block exemption, which are kept under periodic review.

types of data processing activities - and early-stage risk-management (in the spirit of privacy-, data protection-, and security- 'by design').

- Viewed in this context of existing and future compliance requirements which data controllers must already (or soon, come 25 May 2018) heed, it is also hoped to alleviate the strength of potential criticisms that the emphasis on self-assessment under an effects-based approach would increase business costs (e.g. in terms of time and resources needed to be expended on self-assessment compliance exercises). Said otherwise, it can be countered that Model 2 does not import obligations of a type *materially* different over and above what is already required in law (in particular, under the GDPR). Privacy/data protection harm-risk-assessments should already be regarded as a fixed part of the organisation's risk-management strategy. Data controllers should already be considering the specificities of intended data flows where they relate to persons to identify potential data protection issues, including: who collects what information from whom for what purpose; how would the organisation use the collected information; and, what intentions exist regarding data's re-use.[679] Indeed, controllers are already required to inform data subjects about their purposes for processing personal data upon first collection and in

[679] The ICO has pointed out (2017, #2, p.103), in the context of conducting PIAs/DPIAs for big data analytics, that organisations should complete a step whereby they "*describe the information flows"*, and complete *"a systematic description of the envisaged processing operations and the purposes of the processing*…" While the ICO notes the challenges this step can present for organisations, it says (pp.103-104) that these must be faced: "*[d]iscussions with organisations highlighted this step as difficult to complete in the context of conducting PIAs for big data analytics. The consensus was that describing information flows is often much harder because the discovery phase of big data analytics (thinking with data) involves finding unexpected correlations as opposed to the testing of a particular set of hypotheses. Additionally, companies in insurance and telecoms highlighted the difficulty of mapping information flows…It is clear that this step can be challenging for big data analytics, but under the GDPR it will be an explicit part of a DPIA. Furthermore, if the 'legitimate interests' condition is being relied on for the processing of personal data, the GDPR requires it to be described as a part of this step. This requirement links with the new accountability principle in the GDPR which, among other things, obliges organisations to maintain internal records of their processing activities. Therefore, if it's a realistic outcome of a big data project that decisions will significantly affect individuals, every effort needs to be made to observe the requirements of this step by describing the relevant information flows, the purposes for the processing and, where necessary, the organisation's legitimate interests. Although our discussions with organisations revealed a common theme of difficulties with this step, several companies in the telecoms sector emphasised the need for clarity in the aims of data processing and the importance of having an end product in mind. This view is reflected in a paper by the Information Accountability Foundation, which refers to big data analytics beginning with a "sense of purpose" as opposed to a hypothesis. We encourage organisations undertaking big data analytics to think carefully about their sense of purpose for a given project, even if it may change somewhat as the project develops. This will help illuminate the potential information flows that could arise as a big data project progresses. It also complements the advice in our PIA COP about Agile project management and the description of information flows: "Describe the information flows as part of a user story which you can refer to while implementing the project. As the project progresses, record how each stage has changed how you use personal information." For big data projects where there are genuinely no aims or objectives at all at the outset, a potential solution may be to take the processing outside the data protection sphere by using only anonymised datasets during the discovery phase. Should correlations of any interest be discovered, the organisation would then be able to identify the aims of any further processing before starting any analysis of the original dataset containing personal data. At this point, the organisation should therefore be able to describe the envisaged information flows, the purposes for processing and, where necessary, its legitimate interests".* In summary, as a checklist, *"[w]here possible, we clearly describe the predicted information flows for our big data project. If the purposes of the processing are uncertain: we use only anonymised data, or we describe the information flows as the project progresses"*.

> respect of further re-use. Moreover, going beyond information to be given to data subjects, all data controllers should at least to some extent already document their processing activities in order to further transparency and accountability. Documentation is an indispensable internal tool for controllers to manage accountability effectively, as well as for ex-post control by DPAs, and is a requirement under Article 30 GDPR for *all* organisations (without exception) where they are processing personal data "*likely to result in a risk to the rights and freedoms of data subjects"* and *"the processing is not occasional*".

In conclusion, it appears that many of the practical disadvantages associated with the Effects-based Approach to the concept of personal data can be ameliorated, albeit not entirely avoided, through considering insights from competition law, its effects-based theories of harm, and procedural tools that evolved post-modernisation in the new effects-based driven era of EU competition policy. This conclusion is not diminished by the fact that there are outstanding challenges needing to be addressed if the legal concept of personal data moved towards an effects-based conceptualisation. It is deemed premature to analyse these challenges further in this thesis, and they are left outstanding for further research.

In the final chapter, a wider stance (assimilating these conclusions) is taken in concluding on the ultimate research question initially set out in Chapter 1:

> **Would replacing the identificatory requirement for data to be deemed 'personal' under EU data protection law with a definitional provision linked to the assessment of effects flowing from the processing of data better comport with the twin data protection goals of facilitating the free flow of personal data, and safeguarding individual rights?**

Addressing this question ultimately requires consideration of whether the preferred Model 2 is better than *any* iteration of an identificatory-approach model of personal data put forward by others (including a dynamic, risk-based model). More important, why would this be (taking into account factors such as conceptual coherence and robustness, as well as practical implications)? Consideration is also given to how further research paths could be developed in relation to Model 2 in the future as necessary to ensure a strong data economy can be generated while security a high degree of data protection.

Chapter 7 - Conclusions

We return to Chapter 1's discussion and the challenges drawn from the problems described there, followed by a review of the research direction so far navigated.

In 2010, the European Commission stated, "*[t]he concept of 'personal data' is one of the key concepts for the protection of individuals by the current EU data protection instruments [triggering] the application of the obligations incumbent upon data controllers and data processors*".[680] Yet plaguing the scope of associated legislative definitions (under the DPD/DPA/GDPR) are problems of interpretation linked to the identificatory element – i.e. around determining what *exactly* is required for data to be deemed personal beyond it 'relating to' (generally being about) a particular living individual.

Why does this matter? Overly-narrow interpretations of the identificatory element risk ignoring increasingly sophisticated means of re-identifying previously personal data sets, undercutting the protection of privacy and other rights/interests associated with the processing of such data. For example, processing of data relating to persons outside the data protection regime sidesteps supervision by DPAs/courts in case of controller abuse-of-powers. Conversely, overly-broad interpretations of personal data risk expanding data protection coverage to the point of losing practical value. While the EU legislator chose a broad definition of personal data in the 1995 DPD - in light of advantages in "*providing a high degree of flexibility and the possibility to adapt to various situations and future developments affecting fundamental rights*"[681] – recent trends indicate ever-burgeoning interpretations of the identificatory element. Specifically, some authority exists to interpret data protection law as governing the processing of *any* data relating to a person with individual-level elements.

There are conflicting cases/guidance about when data controllers should comply with the obligations imposed by the DPD, and when individuals enjoy data protection rights, regarding similar processing situations across the EU. Specifically, "*the issue of objects and items ("things") linked to individuals, such as IP addresses, unique RFID numbers, digital pictures, geo-location data and telephone numbers, has been dealt with differently among Member States".*[682] This fact led the

[680] European Commission (2010 Communication, p.5).
[681] European Commission (2012, p.14).
[682] Annexes to European Commission (2012, p.6). The quote goes on as follows (pp.6-7): "*[f]or instance IP addresses, which identify computers on networks, are considered as personal data by some Member States, while by others they may be qualified as such only under certain circumstances. Only a few Member States have taken a clear regulatory approach assessing the status of IP addresses. Austria considers IP addresses as being personal data in the Austrian Security Policy Act. Laws in Cyprus, Italy and Luxembourg suggest the same, but within the context of electronic communications. According to the Bulgarian and Estonian Electronic Communications Acts, only a combined set of data*

ICO to remark in 2010 (per the critique in previous Chapters), *"the 'personal data' definition needs to be clearer and more relevant to modern technologies and the practical realities of processing personal data held in both automated and manual filing systems".*[683] The ICO also commented, "*an effective new data protection framework must: be clear in its scope, particularly in the context of new forms of individual identification"* and *"any future framework must deal better with the new forms of identification that are coming into being all the time, particularly in the online environment".*[684]

This thesis contends that the GDPR-articulated identificatory element (like the DPD) does not meet these challenges satisfactorily. Specifically, it does not stymie pre-existing significant

---

which includes IP addresses constitutes, as a whole, personal data. Hence, public authorities in charge of Network and Information Security and Critical Information Infrastructure Protection as well as Computer Security Incident Response Teams (CSIRTs), Internet Service Providers and the security industry have expressed concerns about legal uncertainty regarding the handling and exchange of IP addresses and e-mail addresses across organisations and borders to ensure the overall security of networks and information systems (e.g. to mitigate spam, botnets or Distributed Denial of Service attacks). In the absence of clear regulatory provisions, many national Data Protection Authorities (DPAs) provided guidelines and opinions on the matter. Some of them took the view that the processing of IP addresses does not fall within the scope of legislation implementing the Directive, as long as the addresses themselves are not linked to individuals or to PCs of individuals (e.g. Belgium, UK). The majority of DPAs point to the fact that sophisticated means allow, in most cases, the re-identification of users, and consider, in their opinions on this issue, that IP addresses themselves are personal data (e.g. Denmark, France, Germany, Hungary, Latvia, Lithuania, Netherlands, Poland, Spain). Estonian, Slovenian and Swedish DPAs state that IP addresses are considered as personal data in combination with other data, which could allow linking a dynamic or static IP address to an individual subscriber. The Austrian DPA recognised dynamic IP addresses (which are assigned automatically, as opposed to static IP addresses) as personal data. National courts tend to consider IP data as personal data (e.g. in Austria, France, Germany, Italy, Poland, Spain, Sweden, UK); only few courts found that IP addresses were not personal data since they allowed identification of a computer but not its user (e.g. some courts in France, Ireland). ECJ case law on the confidentiality of electronic communications does not refer to the status of IP addresses. Another major area of divergent interpretation relates to the circumstances in which data subjects can be said to be "identifiable", if they have been made "anonymous", so that data can no longer be related to the individual, or "pseudonymised", where data can only be linked to the individual if one is in possession of a decoding "key". In this regard, recital 26 of the Directive states that "the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable". However, the assessment whether the data allow re-identification depends on the circumstances, available means and technological development. In several Member States, DPAs consider encoded or pseudonymised data as identifiable – and thus as personal data – in relation to the actors who have means (the "key") for re-identifying the data, but not in relation to other persons or entities (e.g. Austria, Germany, Greece, Ireland, Luxembourg, Netherlands, Portugal, UK). In other Member States all data which can be linked to an individual are regarded as "personal", even if the data are processed by someone who has no means for such re-identification (e.g. Denmark, Finland, France, Italy, Spain, Sweden). However, DPAs in those Member States are generally less demanding with regard to the processing of data that are not immediately identifiable, taking into account the likelihood of the data subject being identified as well as the nature of the data."

[683] IC (2010, p.5).

[684] IC (2009, pp.1-2). The IC went on to point out (p.2): "*[i]t is clear that information such as IP logs held by search engines are being used to identify individuals and to take action affecting them, in contexts ranging from behavioural advertising to digital rights management or national security. It is clear that data protection ought to apply to this sort of information. However, we have to be realistic about how such information is treated under the law and what standards we expect those processing it to reach. Whilst we may want this information to be kept secure and protected from inappropriate disclosure, it may be impossible in practice to grant conventional subject access to it or to expect individuals to consent to its processing. The Commissioner hopes that a future framework will treat this sort of information more realistically, perhaps recognising that a simple 'all or nothing' approach to data protection coverage no longer suffices, given the breadth of information now falling within the definition of personal data".* See also similar comments by the ICO regarding the impact of new technologies in IC (2011).

interpretational variances – and exacerbates new potential for disagreement - in theory and practice across the EU.[685] For example:

- The so-called 'data accretion' problem remains: organisations still need to consider the prospect of information they hold about a particular individual becoming identifiable (of that individual) in the event that other data somewhere 'comes into play', or because of information accumulation across different parts of the same organisation.[686] Conversely, data may become more distant from original identifiable records for various reasons. In both cases, it may be difficult to determine at what point data relating to persons becomes (or ceases to be) personal data the processing of which would make it subject to (or no longer subject to) EU data protection law's remit.

- A problem exacerbated regards organisations alleging that information they hold is not personal data (related to the ambit of, what has been termed, the 'non-personal data' concept). For example, the GDPR introduces new obligations and restrictions in association with certain automatic decision-making processing, as well as profiling activities more generally (see e.g. Articles 13(2)(f), 14(2)(g), 15(1)(h), 22, 35(3)(a), 47(2)(e)), but they only apply where personal data are processed. Therefore, organisations that carry out such activities on data relating to persons are incentivised to argue that that the identificatory element is not present in the data they hold to avoid legal compliance obligations. Further exacerbating this problem is the fact that the onus of demonstrating that data are personal data, legally, ultimately rests with DPAs and courts if they want to change/deter behaviour by, say, imposing a penalty - or, uphold the law by finding against a defendant and ordering them to pay damages - for a data protection breach.

- Arguments around the application of anonymisation techniques remain relevant under the GDPR in other ways. Indeed, their use is likely to continue to be relied upon by organisations processing personal data they want to share with external parties for secondary purposes entailing finding untapped data value.[687] Indeed, the future challenges associated with their use will only become harder with the development of the growth of networked objects,

[685] These include ambiguities in the definitions of personal data in national legislation, as well as national differences in interpreting the non-binding but highly influential interpretative Recital 26 DPD in determining whether a person is identifiable (which does not bode well for consistency across EU MSs in interpreting the similarly worded Recital 26 GDPR from 25 May 2018).

[686] In reminder, taking web search as an example, the underlying technological system of an organisation may accumulate search logs containing information about individuals which, in combination, may reach a tipping point whereby a particular person becomes identifiable to it from data (because of this information accumulation).

[687] For example, consider a large database of records indexed by MAC address. This is the sort of data being collected by retailers with in-store mobile phone tracking systems, and used to study how customers move through stores and interact with merchandise.

> involving the collection of very large datasets related to people via sensors constantly exchanging information, and the improvement of analytical techniques). These facts reaffirm the importance of getting clear answers to questions about when information is de-identified sufficiently to fall outside the data protection regime and the relationship between anonymisation and pseudonymisation.

Against this background, therefore, an identification-based approach to the personal data concept in law remains contentious. It begs a general question about what is a 'reasonable' level of personal data legal protection (a question not uniquely EU-centric, being similarly considered in other jurisdictions, per illustrative references to US and Australian legislation).[688] It also gave rise to the thesis research question regarding the determination of whether a different (non-identificatory) personal data model is preferable for better achieving a limitational functionality that is proportionate/fair, while also aligned with policy objectives underpinning data protection law. To assess this proposition, a yardstick of comparison was proposed of 'most likely to achieve data protection's core twin objectives (the safeguarding of rights, and the free EU flow of personal data (in turn requiring a high level of legal certainty))', with a mutually-compatible high level of both adjudged the best regulatory outcome. The related research hypothesis is that a definitional provision linked to effects-assessment – considering the impact of information processing activities upon individuals to whom it relates, rather than identificatory capacity from data - might represent a better proposition under EU law compared to the current and GDPR status quo.[689] This is novel insofar scholarly and policy discussions have not examined explicitly this hypothesis.

Extrapolating this thesis' hypothesis to further unpack answers to the research questions, necessitated developing a model first originated in the proposal for the Effects-based Approach and, thereafter discussed, in respect of models by which it could be evolved. Specifically, in introducing sub-research question 3, alternate 'solutions' were considered from a practical viewpoint involving safe harbour provisions. In turn, this analysis permitted more in-depth comparison between the benefits of effects-based versus identification-based models of personal data, enabling a more-rounded overall answer to the research question.

In terms of remaining chapter structure:

---

[688] See, e.g., in relation to new data protection law recently introduced in Singapore, an article by Lei (2015, Singapore and UK researchers investigate privacy in big data era. [online]), quoting this thesis author.

[689] The wider question raised is whether EU law should avoid a rigid distinction between personal and non-personal data based on identificatory criteria unless it is linked to the circumstances of data processing usage. Analysis in Chapter 3 has illustrated that, under some interpretations of identificatory concepts, there have been attempts to build such links but they strain uncomfortably the identificatory concepts.

- **Section 7.1** concludes on the overall thesis research question, in relation to which the key challenges and findings of the previous chapters are assessed regarding which approach (identification or effects-based) is better on balance. In considering its external value to the academic field, it also underlines the value of using (some carefully selected) presumptions, exemptions, and safe harbours in the field of EU data protection law.

- **Section 7.2** considers what future research might be appropriate flowing from, and left unaddressed by this thesis due to lack of space – specifically, what future research might entail for further development of an effects-based model of personal data that moves from reliance upon risk-assessment to a more risk-proportionate type of data protection regulation (and regulatory system). Moreover, how would this square with the twin aims of data protection in aiming to strike an appropriate balance them?

7.1 Research value and key research drivers

To answer the research question (and sub-questions), Chapter 1 set out a series of tasks.[690] Some of these tasks also point to the importance of the thesis, its analysis, and its novel contribution to knowledge, as part of its main driving aim - to help inform the construction of a more robust, defensible understanding of personal data as a legal concept shaped in alignment with EU data protection law's underlying interests/provisions.[691] Thus, an in-depth evaluation of one of its elements (and, arguably, the most important one) has theoretical value in demonstrating opportunities for shaping this aspect of the definition in ways more justifiable and practical than currently and under the GDPR. Another aim has been to showcase possible variations upon (evolutions of) a new, legal approach and it doctrinally-coherent merits by comparison with the same.

As well as delving into theoretical concepts and related justifications, also considered has been procedural aspects of the EU data protection regime because of problems associated with under-

[690] These are broadly to provide an overview of the DPD (and the UK DPA, as an example of a MS's implementation of the DPD), and GDPR, alongside dissecting the different elements of the definition of personal data in such legislation. This overview has: identified and evaluated current interpretations of the identificatory requirements in relation to the concept of personal data in those legislative documents; identified an alternative theoretical framework of analysis around the concept of personal data under EU law (as well as outlining possible variations upon how this framework may be formulated through changes in law and related justification); and, explored the assumptions and conceptual legacies underlying both approaches, together with the different arguments that may be said to ground each theory, and assessed each theory in terms of how effectively each may realise and reconcile the twin aims of the DPD (mirrored also in the GDPR).

[691] Compare Booth et al (2004, p.19): *"[w]hile ideal types do not by themselves provide a comprehensive understanding of "personal data", they assist in understanding the key elements and in creating a conceptually coherent definition that is justifiable and practical. Using the ideal types as a tool for the formation of a concept of personal data may aid the transparency, accountability and predictability of any data protection authority's decision making strategy and answer to the question "what are personal data?"*

enforcement in determining jurisdiction as to whether data protection applies. In that respect, the last few chapters have focused on practical matters, related to strengthening legal certainty and regulatory incentives to comply with the law, as well as coherency with new, procedural (and risk-based) tools related to demonstrating accountability under the GDPR, which "*sets forth a new legal regime … based on a complete compliance program companies must demonstrate to fulfil".*[692] Said otherwise, taken into account is the fact that *"[w]e're all going to have to change how we think about data protection"* under the GDPR when it comes to organisational accountability, albeit that there are similarities with the existing regime.[693]

In highlighting the potential benefits of the introduction of an effects-based (in particular, a new exemption-centric model), attention refocuses on the importance of safeguarding individuals' interests when data relating to them is intended for processing (i.e. ensuring that they do not suffer appreciable harm) but also recognising the positive effects of associated data processing flows generally. Such analysis could prompt reflection about the basis upon which to devise and implement a modified data protection regulatory regime in the future. As mentioned in Chapter 1, the implementation of the IoT concept – networked objects directly monitoring our physical behaviour – will expand dramatically and with it the amount of machine-generated data relating to us for collection, analysis, and sharing for secondary purposes. In the next decade and beyond, connectivity is predicted to become a standard feature, with billions of Internet-connected objects ranging from telemedicine, to smart meters, to a whole range of new stationary and mobile devices for enabling smart cities, as well as machine-to-machine connected vehicles. It brings with it the prospect for significant benefits and growth. Information from a wide variety of semi-structured data sets will become available from multiple sources for linking and extracting valuable insights in an automated way. However, the privacy challenges associated with mitigating (re-)identification risk will become harder as exploratory analyses on datasets relating to persons continuously churn out hidden (random or non-obvious) correlations/patterns leading to personal revelations yet unimagined.[694] For example, data may be immediately observable and granular, tracking the behaviour/movements of individuals in new ways (including not just device users, but also others in the vicinities of such devices).

---

[692] As Imperiali (2012, p.288) points out, the GDPR also demands a wider *"cultural leap"* within organisations, from a more passive approach of *"formalistic compliance"* to a more pro-active *"policy of daily conformity behavior"*.

[693] See a speech by the UK Information Commissioner, Elizabeth Denham (GDPR and Accountability. London, 17 January 2017, [online]).

[694] Compare, Bridges (2015, p.31): "*[w]ith the collection of very large data sets and the improvement of analytical techniques (the so-called age of big data), the successful de-identification of data subjects has become a very difficult challenge. This challenge – which also may be described in terms of preventing reidentification – will only become harder with the growth of networked objects (the Internet of Things)".*

The remainder of this section highlights the best facets upon which it is concluded to model the legal concept of personal data it being "*any information relating to a natural person*", and evolve the discussion to wider issues around imposing meaningful boundaries around data protection jurisdiction and processing legitimacy, in drawing conclusions.

7.1.1 Dynamic notions over static ones

The analysis carried out in relation to sub-research question 1 in Chapter 3 and Part I of Chapter 5 led to the conclusion that a legal definition of personal data must have sufficient flexibility to allow for future technological developments. However, to maintain that flexibility it should also be context-contingent (dynamic) upon application to the facts at hand, otherwise risk incoherency.[695]

In this vein, the Means Test – and its reliance on reasonableness as a criterion for assessing the likely usage of identification means in respect of specific data (considered in light of associated data environments) - is an important constraint on an exceedingly broad definition of personal data based on the identificatory capacity from data. However, uncertainty persists over how to execute this test as it was intended (such as regarding the extent to which a range of possible factors should be considered and weight applied to them). For example, what if an organisation was able to link additional identifying information after incurring reasonable expenditure, but it was not commercially-useful/practical for it to do so?[696] What should happen if identification was possible, but only using illegal means?

Uncertainty over how to apply the Means Test stokes a risk-averse culture in which, e.g. researchers are inclined to obtain individual consent even for secondary data uses with very low risks of harm for research participants.[697] Moreover, because of confusion over the correct perspective of identifiability to use in applying the Means Test, some may consider data encrypted to industry standards always to remain personal data as long as someone holds the decryption key.[698]

---

[695] Although, to note, Booth et al (2004, p.18) have argued – in discussing whether an effects-dependent model of personal data could be context-independent, *"[t]he need to recognise the context shows the difficulty of maintaining a coherent context independent type".* Notwithstanding, the authors go on to admit (p.19) that, "*Ignoring the role of context causes problems of coherence, but relying simply upon context leads to unpredictability".*

[696] Compare, e.g. Australian LRC (2008, #2, vol.1, p.301) quoting Microsoft Asia Pacific, Submission PR 463, 12 December 2007: "*Just because an organisation holds, or is capable of accessing, various pieces of information about an individual, it can be argued that it does not follow that it will always combine this information to identify them. In many cases it may not be practical or useful for this to be done, and so it simply does not occur*".

[697] RCUK (2010, p.5): "*[a]nother example is large-scale data processing of de-identified data: such data are rich enough that it would be difficult to prove that all records are not personal data (as a small number may indeed be identifiable with sufficient work); however, it is impossible to know which specific records may be high-risk. It is the case that the individual contribution from each record is low (as power of investigation is from large numbers); however, the cost of explicit consent is comparatively high".*

[698] See, e.g. Hon et al (2011, p.46) who propose - in relation to making data protection law fit for purpose for the future – consideration of "*providing explicitly that data encrypted and secured to industry standards (including on key management) are not 'personal data' and thus may be processed' freely by those who do not hold the decryption key*".

This thesis instead focuses on the risk of harm from data processing (surpassing a fixation on whether someone can be identified and/or singled out from information). Chapter 5, for example, contends that assessing/mitigating the risk of appreciable harm to individuals from data in respect of particular data processing activities should become standard practice used by those wishing to modify personal data. This focus is consistent with an emergent trend under the DPD/DPA/GDPR, encouraging and sometimes mandating the use of PIAs/DPIAs both to help organisations decide whether it is appropriate to engage in data uses and to identity alternative approaches that could reduce anticipated impact. Other GDPR provisions, moreover, encourage/mandate organisational accountability through implementing safeguards to identify/reduce the risk of harm presented by future personal data processing activities (e.g. pseudonymisation).

Notwithstanding, as pointed out in Chapters 4-5, context-dependent assessment of harm risk flowing from future data processing activities is also associable with legal uncertainty/unpredictability.[699] Not least, it *"depends upon a reliable prediction of the effect of particular information upon individuals' privacy and this is difficult to achieve in advance".*[700]

Overall, however, it has been argued that in terms of effectiveness – and bearing in mind criticism similarities applicable to both approaches to personal data (especially identificatory models based upon dynamic concepts of identification risk-assessment) – a jurisdictional model based on an effects-based approach is better than one based on an identificatory approach. Yet, it is acknowledged that guidance is required to aid clarity[701] and promote understanding around applying an effects-based model of assessment for jurisdiction-determining purposes. It could indicate how such a model operates in specific contexts, alongside discussing key contextual factors

---

[699] See, e.g. Booth et al (2004, p.19): *"[i]gnoring the role of context causes problems of coherence, but relying simply upon context leads to unpredictability".* As mentioned in Chapter 6, this appears true of any law in the form of general principles, which creates the potential for a significant margin for interpretation and implementation. Compare Cronk (2016, How your legal background may work against you. [online]): *[w]hile the law provides clarity in some situations, it cannot be specific or detailed enough to address every instance, permutation, or nuance of context; it cannot evolve fast enough to keep up with changing social mores and technological innovation; and, if history is a guide, it cannot remain untainted or unbiased enough to properly balance competing interests".* However, this short-failing is particularly relevant to laws that are based on general principles. See, e.g. Australian LRC (2008, #2, Volume 1, p.309), quoting ex Australian Privacy Commissioner (Crompton, 2002): "*[p]rivacy laws need to be in the form of general principles, as information handling is highly contextual. This can create a significant margin for interpretation and implementation…Because of this, elements of the definition of 'personal information' will continue to give rise to theoretical uncertainty".*

[700] See again, e.g. Booth et al (2004, p.18). This difficulty is linked to the fact that per Cronk (2016, How your legal background may work against you. [online]): "*[t]he law can provide a floor by which exploitation of the imbalance is not permitted, but it doesn't achieve a balance of interests. Even standard privacy principles – such as notice, choice, consent, transparency, and even proportionality – are insufficient mechanisms to balance interests. Individuals not only don't know, but in many circumstances, they can't know the full extent of the risks of their disclosures or actions. They simply don't have the time to fully investigate all possible ramifications of their decisions and make rational choices".*

[701] Article 40 GDPR*: "Codes of conduct 1.The Member States, the supervisory authorities, the Board and the Commission shall encourage the drawing up of codes of conduct intended to contribute to the proper application of this Regulation, taking account of the specific features of the various processing sectors and the specific needs of micro, small and medium-sized enterprises".*

to consider in traditional areas of uncertainty (such as scenario-driven processing activities typically applied to dynamic IP addresses).

Notwithstanding, pragmatically, perhaps the theoretical exercise of forcing a direct choice between an identificatory approach and an effects-one is inappropriate. For example, one conclusion drawn about the Effects-based Approach as developed in Appendix 4 is that assessing the likelihood of re-identification occurring from data that has been subject to anonymisation practices is often essential to analysing the *likelihood* that harm will result to the data subject flowing from an intended data processing activity.[702] To such extent, eradicating an identity narrative completely when assessing the non-personal data concept under an effects-based approach seems impossible, albeit that the latter surpasses the former in focusing on quantification of harm magnitude (and its possible mitigation).

Similarly, it appears counterproductive on balance to get rid of all presumptions in this legal area, at least when they have been reviewed and deemed still to retain real aid-to-decision-making value. For example, retaining the concept of sensitive personal data – as indicative of potential harm impact that typically accompanies the processing of such types of data – has appeal taking into account false negatives. Therefore, a composite model drawing on ideas/themes from both approaches may best help decision-making strategies to be more robust. A composite approach would also lessen the extent of any change felt if introduced as it ties in with the GDPR legal framework but refines/extends it.

For these reasons, and as acknowledged in Chapter 5 based on the fact that the (re-)identificatory-approach retained in the GDPR is unlikely to be changed any time soon, it is tantamount to accepting that in the future all data may be deemed personal and could lead to gross-expansion of the legal framework, unless an effects-based narrative is introduced as a binding constraint. In this context, the possibility of introducing a framework of block exemption carve-outs as discussed in Chapter 6 achieves notable value.

It is also undeniable that the identificatory element has at times proven to be a useful tool and retains some value as a rough proxy for whether information relating to persons falls inside/outside

[702] As pointed out, it has also been the preferred option of some scholars (to accept a 'half-way house' model) with a focus on ex-ante risk of harm reduction in practice. See, e.g. Hon et al (2011, p.46): *"[i]t is time to make the DPD fit for the next decade, and hopefully beyond…3. Anonymised data – consider clarifying the circumstances in which anonymisation may produce non-'personal data', with the likelihood of identification being the main determinant - for example treating as 'personal data' data which 'more likely than not' would lead to the identification of individuals. Data should not be treated as 'personal data' where there is insufficient realistic risk of identification. 4. Accountability – consider moving to a more nuanced, proportionate and flexible regime, which bases the general approach on end-to-end accountability rather than the binary 'controller/processor' distinction, basing the applicability of data protection obligations on the risk of harm and its likely severity, with appropriate exemptions. 5. Sensitive data – consider a similar risk of harm approach, with definitions as suggested by the UK ICO."*

the definition of personal data under data protection law. Similarly, it could remain a useful tool (a rule of thumb) for consideration (alongside some other rules of thumbs, such as the motivated intruder test), even if the unlikely were to happen and it were removed from legislation in the future. Indeed, an identificatory analysis or an effects-based one for determining the existence of personal data may often – *although not always* - result in the same outcome in reality as seen in the case scenario discussions in Chapters 4 and 5 albeit the rationale for the outcome is more persuasive if it can be aligned to effects-based analysis.[703] However, the case scenario discussed in the last chapter illustrates that Model 2's non-binary approach can provide additional benefits overcoming existing hindrances to data sharing to start-ups/SMEs for secondary generalised analytics (algorithm creating) purposes with likely positive social effects.

7.1.2 Moving from a binary to a modulated approach that shifts the evidential and regulatory onus of proof

Re-identification technique advances - alongside the acknowledgement of a fluid identifiability spectrum in reality - undermine justifications for applying data protection law in a binary ('all or nothing') fashion. As Polentsky/Tene argue regarding concerns about getting lost in the semantics of this debate even if the (emphasis added), "*a bi-polar approach based on **labelling information** either personally identifiable or not … leads to an inefficient arms race between de-identifiers and re-identifiers. In this process, the integrity, accuracy, and value of the data may be degraded or lost, together with some of its potential societal benefits*".[704] Per Chapter 4, and like the ideas in this thesis, this suggest that the focus of the legal narrative should be on highlighting the essential issues underpinning the tussle of words in labelling information as personal or not to suit their end, whereas the reality may be one that is more modulated and can be envisioned in a non-binary fashion. Moreover, as mentioned, with the introduction of new GDPR profiling rules, incentives increase for online behavioural advertisers to deny that the data relating to persons that they process are personal data (because of the onerous obligations that might otherwise apply to them).[705]

---

[703] This may even be the case in some legal 'grey areas' as the probability that data becomes identifiable (in a dynamic identification-risk-based sense) depends on a range of variables that are a lot like factors that may be considered when assessing likely effects related to the surrounding data environment.

[704] Polonetsky & Tene (2013, p.258).

[705] As mentioned, there is a regulatory gap created whereby organisations can deny that they are processing data relating to persons relying on the argument that direct identifiers (only) have been removed from such data. This type of argument (unless refuted through enforcement action, or in the courts) could result in profiling provisions under the GDPR also being side-stepped) as they only apply to personal data so profiled. The incentive for such denial relates to the fact that – where profiling is carried out based on a data subject's explicit consent – gaining this consent will be extremely challenging especially in the context of online behaviour advertising, where personal data can be used to produce effects on a number of third-party websites. However, profiling (or any other types of automated decision-making) which produces legal effects concerning a data subject - or similarly significantly affects them - can only be

Against this factual backdrop, Chapter 5 acknowledged the importance of data controllers implementing and keeping under review – technical/organisational/legal – measures to protect the subjects of de-identified data as a merit of not just an effects-based approach. Some (re-)identificatory models of a dynamic character also recognise the value of practical risk mitigation by promoting the on-going application of safeguards inhibiting the linking of de-identified data with already identified data. Said otherwise, both approaches can promote reliance upon a comprehensive range of data protection measures, strengthening EU rights' protection. This outcome is better than reliance on models that embed assumptions about the status of pseudonymous/anonymous data as reified states fixed on one side or other of the binary distinction.

Chapter 6 subsequently suggested issues for consideration around potential revision of the GDPR,[706] including introducing a methodology linked to the assessment of the likely harmful effects on data subjects from a processing activity and the mitigation of such effects where they are likely appreciable.[707] Thereafter, further changes were considered by way of softening (modulating) the binary distinction, including a new starting-position legal presumption that all data relating to persons is personal data,[708] but with opportunities to override that legal presumption by obtaining the benefit of a new EU ('de minimis') block exemption regulation. Specifically, a formal legal instrument of this type could be designed to offer exemptions from specific data protection obligations (in respect of personal data that are processed) where the likely harm to privacy resulting from a processing activity is assessed as non-appreciable; effectively, the regulatory burden would be lighter where the associated harm risks are low and yet rights also ensured a high level of protection.[709]

---

justified alternatively under the GDPR (Article 22(2)) where the decision is necessary for entering into or performing a contract, or "*is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests*". In other words, controllers will not be able to rely on their 'legitimate interests' legal basis to justify processing personal data in this context, as they currently can under the DPD.

[706] Specifically, it was proposed that the identificatory requirement in primary legislation should be removed; and, the other criteria for the definition of personal data should remain untouched. The most notable of other criteria being the "relating to" criteria, to ensure that data retains an association with a data subject for it to be subject to data protection rules (with that criteria being interpreted broadly and flexibly).

[707] Comparisons can be drawn here with the Anonymisation Decision-Making Framework (Elliot et al (2016)), which recognises the importance of framing guidance for organisations around the practical application of data protection rules, in particular in relation to risk mitigation.

[708] This suggestion was justified on the basis of the argument that all processing of data relating to persons should be *assumed* (at least prima facie) to merit protection for individuals because it all has the potential for some negative effect on them related to their privacy interests.

[709] Compare by analogy, e.g., the logic of the following argument by Schwartz & Solove (2013, #1, p.913): "*PII 2.0 provides only some of the FIPs to certain kinds of data …PII 2.0 permits variations in protection. Thus, on the surface, PII 2.0 might appear to weaken EU privacy protection and contravene its goal of providing a uniform and high level of privacy protection to data to respect individuals' fundamental right to privacy. In our view, however, PII 2.0 is not only fully compatible with the EU approach, it is consistent with its underlying philosophy and effectively furthers its goals…."*

In practical terms under Model 2, as the proof onus would shift to organisations having to satisfy DPAs that the personal data they control do not merit being subject to all data protection obligations (in relation to particular processing activities), this might also have legal certainty advantages. As long as they satisfy the block exemption conditions achievable by following its methodology recommendations - with the implication that, once followed, the consequences of an activity are not likely to be appreciably harmful for the data subject - they could assume its benefits. Said otherwise, there would become available an automatic exemption from certain legal obligations under EU data protection law. It would be up to organisations to demonstrate methodology fulfilment - in line with a wider GDPR cultural shift to the principle of enhanced-accountability as a means to assure effective data protection,[710] and a more 'hands-on' ex-ante assessment approach to compliance.[711]

On the other hand, Model 2 (and potential theoretical developments thereof) still face challenges as discussed in Chapters 5 (Part II) and 6. Specifically, these relate to the criticism that a modulated exemption approach would deprive data subjects of some of their fundamental rights.[712] Yet, the fact that there are already DPD, as well as GDPR, exemptions stand testament to the fact that, in

---

[710] Bridges (2015, p.36). "*[o]rganizational responsibility for data practices (or what many commonly refer to as "accountability") forms a critical part of effective data protection. Accountability can play a role in responsible information and privacy management practices by offering higher assurances of data protection to individuals and data protection authorities alike...Accountability requires an organization's commitment to and implementation of strong, legitimate and fair information and privacy management practices; an organization's ability to demonstrate the existence and effectiveness of these practices to individuals, regulators, the public, and business partners and internally to management and corporate boards; and an organization's commitment to mitigation and redress for information and privacy management failures. Accountability encompasses not only best practices related to … proper de-identification of personal data, and security breach notification procedures, but also much more. It is both broader and deeper than any single set of best practices".*

[711] Such proposal for offering a safe harbour mechanism for organisations to avail themselves of would not exclude the possibility of ex-post checks by relevant DPAs. Indeed, there could also be introduced a new offence under data protection law – that of, *significantly* harming data subjects from personal data processing. The assumption would be that compliance with data protection rules, which would normally involve mitigating the significant privacy harming effects resulting from processing activity to an insignificant level, have not been followed. The existence of this offence would further incentivise firms to ensure they were diligent in carrying out impact assessments, else risk being fined by DPAs. The introduction of such an offence would also complement Article 22 GDPR, as well as potentially tackle some of the criticisms laid against this provision. For example, the ICO (2012, Initial analysis of the European Commission's proposals for a revised data protection legislative framework, 27 February 2012) criticised a predecessor draft of Article 22 GDPR (Article 20 of the European Commission's draft GDPR text (COM/2012/011 final) for failing to reflect that different degree of risk from profiling, and its different effect on individuals, when it takes place in different contexts for different purposes. It also suggested that a more risk-based approach - perhaps linked to a data controller carrying out DPIAs - could provide safeguards that are more effective. The ICO also stresses that the presence of gaps in the exceptions from the prohibition on processing special categories of data should not lead to a prohibition of otherwise unobjectionable processing. In that context, it proposes (para. 45), *"[o]ne practical solution could be to introduce an additional condition for processing special categories of personal data where the processing manifestly does not impact adversely on the privacy of data subjects*". Moreover, there has been criticism that the right not to be subject to profiling measures has been diluted in Article 22 (2)-(4) of the GDPR, which sets out exemptions to this right – including where the data subject has consented to being profiled - resulting in too few safeguards against the negative effects of profiling on data subjects' privacy. Under a new offence, consent could be no justification for a prior processing activity that caused significant privacy harm to a data subject.

[712] See, e.g. WP218 (pp.3-.4) discussing the "*vigorous debates at the European Parliament and at the Council on the applicability of a lighter legal regime for pseudonymous or pseudonymised data considering that because of their perceived less identifiable nature, the privacy risks for data subjects are reduced*", and the WP's criticisms of this approach.

certain categories of cases, lawmakers have recognised that there is less likelihood of the data subject suffering negative effects in these processing scenarios.[713]

A more significant challenge is the fact that implementation of an effects-based approach under data protection law requires a normative taxonomy.[714] To develop this would require exploring problems of a data protection nature through considering typical data usage contexts, and identifying attendant risks of harms (including from a reasonable-person subject viewpoint).[715] Such exercises could initiate the grounding of what could become more comprehensive/coherent recognition of underlying theories of data protection harm.[716]

Even so, is it clear that Model 2 would be materially better practically than a modulated approach based on a dynamic identificatory model of personal data? Returning to an example from the 'grey areas' of legal uncertainty as to the application of data protection law in certain data processing situations discussed in previous chapters - in relation to pseudonymised data - might help resolve this question.

There are examples of attempts to integrate a modulated approach to identification capabilities in relation to the definition of personal data into legal frameworks. For example, the Austrian Data Protection Act 2000 introduced a category of personal data called 'indirectly personal data' *('indirekt personenbezogener dataen'*) referring broadly to data where the identity of the data subject can be retraced, but not by legally-permissible means. There are special rules under such law for processing such indirectly personal data. For example, there is no requirement for notification of processing of indirectly personal data to the Austrian data protection authority; the use of indirectly personal data shall not constitute an infringement of "interests in secrecy" deserving protection; and, certain data protection rights (to information, to rectification/deletion, and to object to processing) can only be exercised insofar as indirectly personal data are not being processed. Another attempt comes from the European Parliament's GDPR legislative resolution,

---

[713] For example, in respect of Article 11 GDPR as described in Chapter 6. Said otherwise, a modulated approach is acknowledged to provide incentives for organisations to deploy de-identification techniques in exchange for exemption from regulatory requirements.

[714] See, e.g. Booth et al (2004, p.18): "*[t]he list of personal data is difficult to create even when "effect" is narrowed to "effect on privacy" as the type depends upon a reliable prediction of the effect of particular information upon individuals' privacy and this is difficult to achieve in advance. By taking into account social context and removing particular contingencies and individual circumstances, and effectively limiting context, the calculation is easier to make".*

[715] See, ibid: "*[d]efining an individual's society and the effects of that society on information and its relationship to the individual's privacy is difficult, and this can stretch the possibility of applying the model".*

[716] Compare the discussion of context by Nissenbaum (2014, pp.22-23) as a benchmark for privacy online conceived in respect of business models or practices: "*[i]nterpreted as the model or practice of a particular business, context is established according to that business's aims and the means it chooses to achieve these aims…According to this understanding, contexts are defined by particular business models, in turn shaping respective information flow practices. Taking Google's comment above as a concrete case-in-point, this interpretation suggests that contexts generated by their business-driven Internet services, for example, shape consumer expectations of privacy, and not the other way around".* Ibid: *"respect for context would amount to adherence to the set of rules or norms developed by and within respective sectors or industries".*

proposing to define 'pseudonymous data' formally and exempt the processing of data matching this definition from certain obligations. However, practical challenges associated with both approaches exist around finding precision legal definitions for these 'new' data categories.

As well as problems around clarifying when modulated rules might apply, the 'why' (explaining the justification) for derogations from certain data protection rules – e.g. in relation to the processing of certain online data – is equally important. It should be linked to the outcomes sought for data subjects in typical data processing scenarios that might affect them. Furthermore, it should include consideration of processing risks that have an especially adverse of discriminatory effects on particular individuals, but also groups of individuals or on society more widely. In turn, reasons given should include the explicit airing of considerations around data protection obligations being proportionate to the risks of harms to those potentially affected and their individual/collective interests.

For example, the ICO has stated that the need for a "*more flexible and contextual approach*" (in relation to modelling a legal definition of data sensitivity – 'special categories of personal data' – using an effects-centric analysis), including suggesting:[717]

> [T]he distinctions between special categories and ordinary data could be removed from the new framework, with emphasis instead on the risk that particular processing poses in particular circumstances. It is important to give a message to data controllers that a simply binary (special categories – the rest) approach is not good enough, and they must consider the context in which they hold information and the risk this poses to individuals. It would be helpful if national data protection authorities or EU-level bodies… could produce guidance with examples that could help organisations to assess genuine sensitivity in various contexts.

Thus, the ICO is pushing broadly to change the legal narrative, as well as expectations, to emphasise assessing processing risks and lowering these where possible, whereas currently the prevailing narrative about personal data is often about encouraging those who hold data relating to persons to hold it in the least identifiable form possible. Although, as noted, in many cases, the two narratives are often confused, e.g. per Solove/Schwartz:

[717] IC (2010, p.3). The ICO also comments in the same document (p.2) in relation to other types of personal data (information such as IP logs), "*[w]hilst we may want this information to be kept secure and protected from inappropriate disclosure, it may be impossible in practice to grant conventional subject access to it or to expect individuals to consent to its processing"*

> [W]hen a breach involves only identifiable data, the harm that the information can cause to individuals is much less likely to occur. Harm can only occur when the party who obtains the data also knows how to identify it.[718]

7.1.3 The importance of the thesis' contributions, in particular for forging a new effects-centric framing narrative to underpin data protection law jurisdictional issues

An effects-based approach is no panacea to the fact-driven problems described in Chapter 1. Partly for that reason, we cannot expect a dynamic definition of personal data to eradicate legal uncertainty associated with its application entirely. However, some areas of identificatory interpretational uncertainty exceed that caused by the contextual aspects of privacy-related concerns.[719] Mixed meanings attributed to terminology such as anonymous/ pseudonymous data, and deciphering the Means Test, compound uncertainty levels and suggest a clearer guide to classificatory decision-making is required to minimise current levels of unpredictability likely enduring under the GDPR.

In this context, an important thesis contribution looked to address a practical organisational issue about how to make judgement calls on whether they are subject to data protection by investigating the rationale for 'grey area' decisions.[720] Inserting an effects-based approach into the definition of personal data would provide a narrative-framing-functionality around harm assessment/mitigation associated with planned data processing activities.[721] Arguably, moreover, this was originally the legislator intention in positing an identification-from-data test as a minimum scoping mechanism in law, because data protection was designed to protect individuals against the harmful impact resulting from the processing of data relating to them. Furthermore, this contribution value will increase as more organisations navigate routes between data protection compliance and processing data relating to persons in ways that avoid stripping out entirely opportunities for data utility extraction, requiring risk-assessments dependent on data processing context.

---

[718] Schwartz & Solove (2013, #1, p.915). Therefore, again, in practice, it is unlikely to be possible to entirely separate a narrative of (de-)identification from one of consequential harm (and harm mitigation).

[719] Moreover, it is not just privacy considerations that are contextual, so too are other risk-assessments that organisations have to engage with, such as ensuring they have adequate security management. All such issues involve organisations in gauging eventualities, making assumptions about possible harms, drawing conclusions that require reasonable arrangements be put in place, and documenting these to evidence that they have been considered, as well as returning to these assessments periodically to review them.

[720] Zwenne (2013): "*clarifying and explaining data protection law is not an easy task. A [data protection authority] that wishes to be credible should do its utmost to explain its reasoning, or at least it shouldn't mind doing so*".

[721] EDPS (2015, p.18). "*ACTION 3 – 'Increasing transparency, user control and accountability in big data processing' - Develop a model for information-handling policies, particularly for online services provided by EU bodies, which explains in simple terms how business processes could affect individuals' rights to privacy and protection of personal data…"*. In turn, this would aid the accountability and predictability of any data protection authority's decision-making strategy.

Doctrinally also, an effects-centric narrative could help develop theories of data protection harm, a job no doubt better suited to non-lawyers in analysing framework factors for determining where regulatory efforts (and resources) would be best suited. The knock-on-effect hope is that identifying the relevant/significant elements justifying a more coherent theory of personal data definition could also lead to more reasoned (and transparent) levels of decision-making and legal certainty.[722] Backed up by best practice methodology (a framework of assessment),[723] it would also recognise the importance of 'soft law' (non-formally-binding) solutions, as ways of sharing best practice with explanations providing a more highly-calibrated view of factors for determining when data protection obligations trigger. This does not preclude possibilities for issuing additional sector-by-sector guidance to fill in the gaps on how a methodology would work in specific' grey area' contexts oft-encountered in specific sectors.[724]

Furthermore, as well as making laws comprehensible/certain, workability (relative ease of regulation/enforcement) is important. Previous analysis (particularly Chapter 6) acknowledged some such considerations, albeit only partially. One example is industry cost – especially, to SMEs – from changes to the law, typically requiring a clear-cut case with compelling justifications.[725]

Beyond achieving data protection's twin aims, there are of course other issues for consideration in justifying some further degree of law reform after GDPR. First, processing that might have little impact on individuals generally may in fact have significant effects on certain groups and society at large. Per Chapter 4 and above, the definition of personal data and its conceptual relationship to the potential for harm to group interests is an overdue challenge that needs tackling. More broadly, the ability to take such collective effects into account is sorely needed as part of the on-going risk management narrative gaining force in modern data protection law discourse. This thesis highlights

---

[722] It could also be seen as an opportunity to bring the EU definition of personal data into more alignment with other non-EU privacy laws, such as in the US and US legislative approaches to PII. In the US, the regulation of privacy has been traditionally viewed more in terms of avoiding harm to people in specific contexts. Compare Schwartz & Solove (2013, #2, p.4): "*PII 2.0 fits with the US harm-based approach because its tiered approach provides more privacy protection when there is a greater risk of harm. This approach contrasts with the EU approach to PII which defines it quite broadly and provides a full suite of rigorous protections to a wide array of data even when there is little risk of harm".* See also Bridges (2015, p.11): "*Our goal is to identify a few selected practical steps to bridge gaps between the existing approaches to data privacy of the EU and US in a way that produces a high level of protection, furthering the interests of individuals and increasing certainty for commercial organizations. These "privacy bridges" are designed to advance strong privacy values in a manner that respects the substantive and procedural differences between the two jurisdiction. This bridge calls on EU and US regulators, who already share common views about de-identification, to identify concrete, shared standards on de-identification practices".*

[723] Compare, e.g. Elliot et al (2016). In other words, per Chapter 6, a methodology should involve consideration of factors such as the broad context of the data environment, including looking at business plans and how they might evolve over time (as with any business analysis) and planning to review periodically (as classifications of data from personal to non-personal –or vice versa – may change over time, as processing plans change).

[724] Such guidance could also potentially include sector-specific block exemptions, which provide industry sectors with flexibility in their interpretive approach to meeting statutory requirements.

[725] Otherwise, the change could result in implementation costs that are disproportionate to any benefit that may be obtained with respect to the protection of individuals' interests. In particular, guidance would be needed to help with the specific needs of SMEs in applying the law.

how critical tools, such as block exemptions and related processes, could help overcome previous informal attempts in this area by providing a structured way to embrace a wider reflection on the impact of processing on a wide range of interests.

An effects-based model, for example, has been argued as potentially more pragmatic as a way of dealing with things such as group privacy (e.g. genetic classes, who could be recognised as holding collective interests on the level of the genetic category) and the potential for significant social or economic disadvantage. It is clearly useful as a tool for thinking about how data protection needs to develop in the future within a structured framework, whereas currently the idea of group privacy has little legal traction. In this sense, there are some similarities with consumer law, where collective interests are recognised (e.g. in relation to product security), but the parties potentially harmed by unfair practices may be unconnected. This issue is a strong candidate for further research, particularly because it may encompass scenarios in which processing of information which is not, and perhaps never has been, personal data (as currently defined) may nevertheless result in substantial harm being caused to groups based on one or more specific characteristics that the group members share. Moreover, data processing actions regarding a particular group could create conclusions with implications for other groups (related genetics groups being the obvious example again here, but also potentially non-apparent algorithmic groups 'seen' using sorting-analytic metrics whereby the aim of the grouping may not be to identify the individuals within it). We have reached a stage in the development of data protection where groups also need certain legal protection from harms as entities in their own right, and this requires a new approach that goes beyond traditional approaches to data protection.

While future success depends on finding better consensus on the harms for individuals and collective interests that data misuse can cause, and an appropriate range of mitigatory measures suitable in the area of data protection, Model 2 is an effective starting point for achieving this goal in a scalable way by creating a workable framework that the DPAs will have direct input into. However, work is still needed in the long term by generating a level of understanding and knowledge about the degree to which processing practices in typical contexts may harm individual and collective interests, as well as how they are likely to maximise the potential benefit from data usage. Part of that conversation will also need to cover distributional issues of data/privacy protection – i.e. how to best distribute the benefits of data use across society while ensuring acceptable fairly prioritised levels of risk for individuals and communities. Future work may wish to investigate how a fair distributional system could be achieved.

Second, Model 2 also provides a more pragmatic way of dealing with another challenge not visualised by existing forms of data protection, viz. problems associated with technologically and

social developments (e.g. smart buildings and smart meters) providing impetus for more and quicker extraction of value from data. It is a compliance tool that can be developed to encourage analysis and managing of the benefits that big data and machine learning can create using objective standards, hand in hand with better data protection by design. Moreover, per the case study discussion in Chapter 6, the potential benefits of Model 2 over the status quo described there could become particularly relevant as objects interact and share personal information. This is because reliance on consent to legitimise frequent and unobvious exchanges of data is likely to become more and more unrealistic and inappropriate. While the legitimate interest test may be available as an alternative basis to justify secondary processing, it is argued that Model 2 is better suited as a simpler system based on objective standards and theories of harm for development. It has the benefit of avoiding a non-scalable and therefore highly intensive regime of obligations and responsibilities under the GDPR (with similarly complex legal documents to cover all bases), which is practically unmanageable. The solution proposed would not eradicate all risk, but should help those considering data sharing arrangements take comfort in determining whether or not the level of risk is acceptable in the circumstances, and minimise risk using scalable protection, taking into account the benefits of what is hoped to be achieved. In other words, it would provide an objective framework to provide data-innovators with more confidence about securing beneficial uses of data that might otherwise be missed.

Notwithstanding, the extent of change to practices associated with the introduction of an effects-based exemption approach capable of responding to new data technologies should not be overestimated. As part of the new GDPR era, organisations will need to perform on-ongoing data protection processing assessments and document these as part of their day-to-day compliance obligations. These include record keeping (Article 30) about processing activity-by-activity: purposes; categories of data involved; categories of recipients to whom personal data is disclosed; erasure time limits; and documenting security measures taken by controllers/processors). Thus, organisations must have processes in place that enable them to engage in these risk-assessments to comply with the GDPR; whereas the carrying out of objective assessments – as evidenced through documentary records and the implementation of safeguarding measures in light of identified risks – are likely to be taken into account as fine-reducing mitigatory factors by DPAs.

While, as has been pointed out, the GDPR broadens the relevance of anticipative risk-based assessments, in particular in the context of referring to risks-to-rights to the fundamental rights and interests of data subjects that should be taken into account (e.g. at Recital 74 and with reference to DPIAs requiring organisations to assess the likelihood and severity of such risks likely arising with certain types of data processing). This fact has led scholars such as Van Dijk et al (2016) to posit a new 'risks-to-rights' approach emergent under the GDPR and, in particular, from its Recitals 75-77

referring to risks to the rights and freedoms of natural persons, risk-assessment, and risk-assessment guidelines, as well as Recital 84 referring to risk evaluation and impact assessment (together with mirroring provisions in the GDPR Articles, e.g. 24(1), 25, 32(2), 33, and 35).

Notwithstanding, while an effects-based approach could be seen as part of this (what might be termed, more accurately, a risks-'*of impacts upon*'-rights) approach, as Chapters 4-6 makes clear the former requires *negative* impact (consequential and concretisable harm, i.e. impact remaining grounded in concrete risks of harm to individuals) resulting to the data subject and is limited to a jurisdictional context.[726] By contrast, the WP has delineated an assessment of risks-to-rights as requiring consideration of when processing of personal data might lead to either *positive or negative* impact-risks to rights conceived abstractly.[727] Furthermore, the WP suggests that assessing risks-to-rights is not underpinned by a fully-formed approach – and single logic - applicable whenever this phrase is used, but implies instead a series of different impact-assessment exercises) as it states that the legal assessment of impact should be understood in relation to the particular legal rules in relation to which it being used (e.g. under Article 7(f) DPD, dealing with determination of whether the legitimate interest legal basis can be relied upon).[728]

Thus, while an effects-based approach can take inspiration from risk-based methodologies (in data protection law and in other fields) – and in particular, from impact-based methodologies – it is intended to be unique insofar as it relates to jurisdictional issues and the satisfaction of exemption conditions in that context. This statement does not preclude consideration of the general logic underlying the choice of key factors adopted by other impact-assessment methodologies (such as

---

[726] An effects-based approach proposition has also tried to address some of the criticism of the risks-to-rights approach outlined by Van Dijk et al. These include, e.g. criticism relating to specification of the criteria for what will constitute the content of the risk, how such risks may be identified, and regarding the notion of probability.

[727] See WP203 (p.25) and WP217 (p. 37). For example, the latter Opinion states: *"[t]he Working Party emphasises that it is crucial to understand that 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…the notion of impact, as used here, encompasses the various ways in which an individual may be affected - positively or negatively - by the processing of his or her personal data".* The WP also admits in WP217 (p.38): *"in applying the methodology, it should be recalled that assessing impact cannot lead to a mechanical and purely quantitative exercise"*, and (pp.53-54): *"[s]uch prior assessment should not be too burdensome, and remains scalable: it may be limited to essential criteria if the impact of the processing on the data subjects is prima facie insignificant, while on the other hand it should be performed more thoroughly if the balance was difficult to achieve and would require for instance adoption of several additional safeguards…Just as it is scalable in how much detail the assessment needs to be carried out, the extent of documentation should also be scalable. With that said, some basic documentation should be available in all but the most trivial cases, independently of the appreciation of the impact of the processing on the individual. It is on the basis of such documentation that the assessment of the controller may be further evaluated and possibly contested".*

[728] WP217, (fn.84): "*[t]his assessment of impact must be understood in the context of Article 7(f). In other words, we do not refer to a 'risk analysis' or a 'data protection impact assessment' in the sense of the proposed Regulation (Articles 33 and 34) and the various LIBE amendments to it. The question what methodology should be followed in a 'risk analysis' or a 'data protection impact assessment' goes beyond the scope of this Opinion. On the other hand, it should be kept in mind that - one way or another - the analysis of impact under Article 7(f) can be an important part of any 'risk assessment' or 'data protection impact assessment' and can also help identify situations where the data protection authority should be consulted".*

the nature, scope, context, and purpose of the processing as key factors described in the GDPR – e.g. Recital 76 - for assessing impact, its likelihood and severity), and how they have been developed into more detailed multi-criteria evaluation tools. In other words, it is important to build upon and take into account current laws, research and guidance on risk, risk-assessments, and impact-based assessments in the field of data protection, including under the GDPR alluding to risky processing activities that may result in harm, as well as on the nature of these harms. At the same time, it is recognised that there is no consensus on the specific types of harm liable to arise – or, indeed, comprehensive and conclusive guidance as yet on how to assess negative impact risk - in a data processing context.

7.2A possible future research agenda

This final section revisits some of the unaddressed points raised by the proposition of adopting an effects-based approach of some kind to jurisdictional issues, in particular Model 2, in terms of pointing out some key limitations of the scope of the current research project, alongside the ways in which future research might develop the thesis.

7.2.1 The personal data concept has multiple inter-linking building block elements

The analysis in this thesis focuses on the identificatory element in the legal definition of personal data. Its remit was not fully extended to consider the other three building blocks of the personal data concept (per Chapter 2). Yet, these four elements should be seen as closely intertwined; especially, the relation between the identificatory element and the 'relating to' element both stand in association with the same object and subject of the definition (the data and the person). Arguably, the dynamic quality of the personal data concept adheres to the transitive aspects of *both* these elements in combination, and yet how the one is meant to interact exactly with the other is not always entirely clear.

7.2.2 Risk-management depends upon consensus on the harm for protection and its quantification, a concern surpassing issues of jurisdiction

As mentioned, risk-management principles/procedures have long reverberated in data protection law.[729] Yet, the field of expertise developed around risk-management is not fixed to law and often resides in non-legal areas (e.g. insurance, health and safety, and environmental risk-management). Moreover, the starting point for effective risk-management (the assessment of likely risks so they

[729] Cate et al (2015).

can be identified/mitigated) typically arises after consensus has been reached on the particular harms in a particular area that it is intended to identify and mitigate.[730]

Contrastingly, a lack of consensus (so far) on this issue in the field of data protection law means that the conceptual foundation upon which an effects-based (or, indeed, any harm-based) approach can be tethered is not yet fully developed. In other words, it is difficult to guard effectively against threats associated with data privacy and anonymity until we have a clearer idea of what data protection law is intended to be guarding against exactly and in what forms. This starting-position state-of-play present opportunities ("*to develop modern, effective risk management tools and a framework of impacts—both harms and benefits—building on decades of experience with risk management broadly*"), but also challenges ("*to do so quickly to keep pace with dramatic changes in technology and human and institutional behaviour*").[731] Said otherwise, developing international consensus around a taxonomy of data protection harms, and a framework for assessing them, goes hand-in-hand with learning from the field of risk-management. As mentioned, it could also pave the way for further evolution of an effects-centric approach under data protection law (e.g. through the introduction of a new offence for significantly harming data subjects from personal data processing – see fn.711).

Thus, it could be a driving force for moving towards a broadly-consistent framing of data protection rules focused on prevention of harm (in particular, to ensure that processing activities may only be carried out on data relating to persons if they carry a non-appreciable risk of harm to data subjects, and – as mentioned above - encompassing consideration also of potential harm to data subject groups and collective interests upon further research of model development in this respect.[732]

This framework would be based upon: the need for assessment of the harm that may be suffered by individuals and collective interests caused by processing activities upon data relating to them; the broad criteria of evaluation used in terms of determining the potential magnitude and likeliness of harm arising from intended processing activities; and, providing incentives for taking remedial actions as necessary in order to avoid or minimise negative impacts. It aligns with the imminent

---

[730] Centre for Information Policy Leadership (2014, p.1): "*[d]ata protection has long relied on risk management as a critical tool for complying with data protection laws and ensuring that data are processed appropriately and the fundamental rights and interests of individuals are protected effectively. Yet these risk management processes, whether undertaken by businesses or regulators, have often been informal, unstructured and failed to take advantage of many of the widely accepted principles and tools of risk management in other areas. In addition, institutional risk management in the field of data protection has suffered from the absence of any consensus on the harms for individuals or negative impacts that risk management is intended to identify and mitigate in the area of data protection. This is the starting point for effective risk assessment in other fields".*

[731] Ibid, p.1.

[732] See e.g. Floridi et al (2016), which contains a series of articles highlighting issues associated with group privacy and new challenges of data technologies to such interests arising in response to specific challenges. See, e.g. comments in WP251 (p.11): *"[p]rocessing 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".*

greater use of upfront impact assessments as an aid to GDPR compliance – alongside more data controller/processor accountability - as well as the push towards strengthened incentives for anonymising/pseudonymising data collection in order to minimise personal data volumes collected.[733]

7.2.3 Towards more risk-proportionate data protection regulation

Chapter 6 introduced questions about the extent to which an effects-based approach to personal data lead us, inexorably, towards exploring further theoretical developments by way of a more whole-hearted acceptance of risk-proportionate data protection regulation.[734] Said otherwise, this thesis and its Model 2 proposition might be considered the start of a much wider research project into the benefits of introducing a more graduated (non-binary) risk-based approach to applying data protection rules using effects-centric exemption models.[735]

Specifically, exploring this potential research direction-of-travel envisages a data protection regime aiming for more proportionality in its rule application: matching the triggering of subsets of legal obligations (viz. the severity of the overall regulatory obligations to be discharged) broadly to the risk level of harm, associated with particular activities.[736] Therefore, the regulatory burden would be lighter in respect of the processing of personal data likely reasonably resulting in a low risk of harm to data subjects (than under the status quo). As commented by Hon et al, in the context of proposing a "*continuum or spectrum of parties (depending on the circumstances) who may be*

[733] Article 29 Working Party, Advice Paper of 13 May 2013.

[734] The adjective "wholeheartedly" is used as there are already examples of a risk-based rule-flexing approach in data protection law as previously mentioned. For example, see WP136 (pp.4-5): "*[f]lexibility Is embedded in the text to provide an appropriate legal response to the circumstances at stake… the exemptions under Article 3 take into account the technical way of processing (in manual non-structured form) and the intention of use (for purely personal or household activities by a natural person). Even where processing of personal data within the scope of the Directive is involved, not all the rules contained therein may be applicable in the particular case… In those cases where a mechanistic application of every single provision of the Directive would at first sight lead to excessively burdensome or perhaps even absurd consequences, it must be first checked 1) whether the situation falls within the scope of the Directive, in particular in accordance to Article 3 thereof; and 2) where it falls within its scope, whether the Directive itself or national legislation adopted pursuant to it do not allow for exemptions or simplifications with regard to particular situations in order to achieve an appropriate legal response while ensuring the protection of the individual's rights and of the interests at stake. It is a better option not to unduly restrict the interpretation of the definition of personal data but rather to note that there is considerable flexibility in the application of the rules to the data…the text of the Directive invites to the development of a policy that combines a wide interpretation of the notion of personal data and* ***an appropriate balance in the application of the Directive's rules"*** (emphasis added).

[735] This is different from risk-based data protection regulation (viz. a more effective regulatory risk allocation so DPAs are able to make decisions based on genuine risk to optimise their effectiveness). It is also not referring to a risk-based approach construed as a valuable tool for calibrating the implementation of and compliance with privacy requirements, prioritising action, raising and informing awareness about risks, identifying appropriate mitigation measures and, in the words of the WP, providing a "*scalable and proportionate approach to compliance*".

[736] Schwartz & Solove (2013, #2, p.4): "*PII 2.0 also builds on evidence that EU is evolving past a simple one-size fits-all privacy protection regime".* See also Hon et al (2011, p.41): "*[t]he criteria for triggering the application of data protection obligations need to be more nuanced, rather than 'all or nothing'. It may be appropriate to apply all the Principles in some situations, but not in others".*

processing personal data, each having varying degrees of obligations and liabilities under data protection law":[737]

> How the information should be processed should then be tailored accordingly, considering what measures are appropriate in the circumstances in light of such risks… rather than applying all the Principles to information which has been determined to be 'personal data', there should be consideration in each particular context of which data protection rules should be applied, and to what extent, based on the realistic risk of harm and its likely severity.[738]

Developing more exemptions from data protection obligations in respect of processing situations that have been assessed low risk of harm is not incompatible with protecting all possible individual's rights/interests at stake. After all, the GDPR recitals state that *"[t]he right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced with other fundamental rights, in accordance with the principle of proportionality*".[739] Moreover, as the WP points out, a number of DPD provisions "*contain a substantial degree of flexibility, so as to strike an appropriate balance between different interests*".[740]

Another way-in for such an approach could be via the explicit recognition in legislation that data controllers must engage in ex-ante harm identification/mitigation analyses of processing activities affecting persons, as well as analyses regarding potential benefits flowing from processing in certain situations. The latter recognition of positive effects in law would be over and above – and distinct from - an implied balancing exercise already embedded in data protection law (derived from the fact that pro-processing factors e.g. related to gaining utility from anonymised data, can be weighed against anti-processing factors e.g. rights to privacy, dignity, identity, self-determination, non-discrimination, etc.). While the use of large data sets for research purposes can directly benefit society highlighting the social utility of data analysis and research, those who provide data for such

---

[737] Hon et al (2011, p.45): "*[a] risk-based approach, with reduced regulation of personal data in certain situations…considering a continuum or spectrum of parties (depending on the circumstances) who may be processing personal data, each having varying degrees of obligations and liabilities under data protection law, with the risk of identification and risk of harm (and its likely severity) being the key factors. Such an approach should result in lighter, or even no, data protection regulation of passive utility infrastructure cloud providers, while reinforcing the obligation of cloud providers who knowingly and actively process personal data to handle such data appropriately".*

[738] Ibid, p. 37.

[739] GDPR, Recital 4

[740] See WP136, p.4: "*[a] number of provisions of the Directive contain a substantial degree of flexibility, so as to strike the* ***appropriate balance between*** *protection of the data subject's rights on the one side, and on the other side the legitimate interests of data controllers, third parties and the public interest which may be present. Some examples of such provisions are contained in Article 6 (retention period depending on data being necessary), 7.f (balance of interest to justify processing), last paragraph of 10 (c) and 11.1 (c) (information to the data subject where necessary to guarantee fair processing), or 18 (exemptions from notification requirements), just to mention a few cases"* (emphasis added).

ends expect a high level of protection against a variety of privacy harms, ranging from inconvenience or embarrassment to identity theft. All of these techniques involve some trade-off between privacy and data utility. The question is how to minimise data risks while maximising benefits to all parties. Per the House of Commons Science and Technology Committee (2016, p.5):

> The anonymisation and re-use of data is becoming an issue that urgently needs to be addressed as big data becomes increasingly a part of our lives. There are arguments on both sides of this issue: Seeking to balance the potential benefits of processing data (some collected many years before and no longer with a clear consent trail) and people's justified privacy concerns will not be straightforward… to strike a transparent and appropriate balance between those benefits and privacy concerns… in general, modifying the data to protect against attribute disclosure means reducing the plausible inferences that can be drawn from the data. This can be detrimental to the objective of learning as much as possible from the data and building generalizable statistical models from the data. Furthermore, to protect against attribute disclosure, one must anticipate all inferences and make data modifications to impede them, which may not be possible. Some inferences may be desirable because they may enhance understanding of the treatment benefits or safety of a new drug or device, and some inferences will be stigmatizing to the data subjects. Legislation should be kept to the minimum required to facilitate the uptake of the Internet of Things. It should enable more efficient public and private services in areas such as healthcare, energy and transport, and should aim to minimise threats and harms.

It is within this narrative that a modulated effects-sensitive exemption model neatly fits, as Hon et al point out:

> A better starting point might be simply to require explicit consideration of the degree of risk of harm to living individuals which the intended processing carries, and the likely severity of that harm, balancing the interests involved, and subject to appropriate exemptions. This would require an accountability-based approach that is proportionate to the circumstances, including the situation of the controller, data subject and any processor. More sensitive situations, with greater risk of resulting harm and/or greater severity of the likely harm, would require more precautions than less sensitive ones. Such an approach would be in line with the European Commission's desire to require additional protections in certain circumstances. Yet it would also allow fewer or even no Principles to be applied, for example, to encrypted data held by someone who has no access to the

key, where the encryption and other security measures have been to accepted industry standards and best practices, or where recognised certifications have been achieved.[741]

7.2.4 Towards a more coherent data protection future legal regime

The GDPR was designed to address the significant evolution in technology since the DPD was introduced. Yet it was decided, in this thesis, to consider both (the DPD to set the scene, and the GDPR as the future 'script' in this area) regarding their fitness for purpose within the scope of the research question(s). However, in considering how to make data protection rules fit for the next twenty years, we should think beyond 2020 about a world in which the IoT seeps into almost every aspect of daily life. How data protection law might evolve to deal with the impact of these emerging challenges needs anticipating now. How can we supersede the currently dominant legal paradigm focus primarily on individual interests and identifiability, things that are increasingly incidental in the big data era?

Having a clear and logically-coherent legal framework for distinguishing between personal data and non-personal data will become increasingly important. It will need to be a flexible model to react to technological change, and balance the consideration of potential benefits/harms associated with regulating this area. Indeed, as argued, it may make sense for there to be a legal presumption that raw data generated by connected technology relating to persons contains personal data because of the extent of this challenge and the risk of false negatives, albeit one that could be effectively rebuttable in one or more respects in terms of the compliance obligations imposed in relation to its processing. A data protection block exemption could fit neatly into such a future and its regulatory landscape for these reasons.

Finally, there is the issue of cross territorial compatibility. Globalisation, alongside technological advances, poses common challenges to developing a progressive, sustainable model for protecting privacy in an international data environment. Taken beyond the context of the EU, these problems amplify because of the lack of consistency on the definition of personal data/PII. Carrying out research into the potential for consolidating different approaches to the concept of personal data could help global trading relationship and rights protection. An effects-based approach could be the start of bridging such divergences.

[741] Hon et al (2011, p. 41).

Appendix 1 – Discussion of identification-related terminology from a non-legal perspective

This Appendix introduces key terms related to identification capabilities, including some terminology introduced in Chapter 1. This discussion is intended to help scope out their broad meanings in illustration of the different types of identification capabilities that can exist and to highlight some ambiguities around meaning as often used in everyday discourse. As Chapter 2 introduces the legal framework and its use of key legal terms in complement to this analysis, this discussion also prepares for it by promoting consideration of – not just what key differences in nuances of interpretations exist around their usage – but also why there exists some terminological overlap giving rise to some confusion and potential inconsistencies in the use of the key terms (such as 'identified' and 'identifiability'). In so doing, it also sets the groundwork for Chapter 3 and its dissection of some of the theoretical assumptions underpinning the legal discussions using these terms.

1. Identifiers

Identifiers are objects holding a close relationship with a person, from which they can be recognised or established as being them. One conceptualisation of identifiers is that they are inherent to an individual's being. Another – a more subjective conceptualisation – is of things that link to a specific person, but are not necessarily inherent to them and are variable as long as they may be used (by someone at a particular point of time) as indicative of that person and not another.

(a) Direct identifiers

Identifiers deemed to have strong links to a particular person are often called 'direct' identifiers. Typically, they may be considered things capable of identifying (see discussion next of the verb 'to identify') someone without any other information required. Notwithstanding, at least in common day speech, they do not necessarily have to be unique. For example, your full name may not be unique compared to the rest of the world's population, but it may still be deemed to have a strong association with you because of enhanced possibilities to relate different pieces of information from different sources to the right you (e.g. if such information is searchable online).[742] In that

[742] For example, direct identifiers have been defined as, "*data that can be used to identify a person without additional information or with cross-linking through other information that is in the public domain*" (ISO/TS 25237:2008).

context, if the forms of direct identifiers include data unique to a particular data context, either intentionally or by association, they may also be potentially be considered to cover data that is functionally unique.[743]

(b) Indirect identifiers

An 'indirect' identifier, by comparison, may be conceived as something with less linkability to an individual and identifying someone from it only possible by combining it with other information. Indeed, the revelatory power of indirect identifiers may be limited to its association with more directly-identifying data (such as IP addresses associated with stored search results, or an email address, in context).[744]

(c) Quasi-identifiers

In reality, going beyond this direct/indirect distinction, all data that can act as an identifier should be considered to exist on an identifiability spectrum (see Section 1.2 above), with its strength-of-personal-association to an individual to which it relates often depending on the relevant factual context and analytic utility, as well as the particular point in time being considered (as the association between person and identifier may not be permanent).

From this perspective, even generic, often analytic-useful, characteristics relating to a person in data - like age and gender - can be indirectly identifying of individuals behind the data when combined with other information resulting in unique (more or less) data property sets.[745] Specifically, the use of advanced analytics on multiple data points about a person can give rise to the possibility for any non-unique attributes to form 'quasi-identifiers' (viz. indirectly identifying

---

[743] Compare Elliot et al (2016, pp.17-18), who describe five types of 'direct identifiers': 'intentional unique identifiers' (e.g. US social service number); 'digitised unique biometrics' (e.g. digitised fingerprints); 'associational unique identifiers' (e.g. mobile phone number); 'transactional unique identifiers' (e.g. browsing-session cookies); and, 'functional unique identifiers' (e.g. full name and address). The authors label the last category as a "*borderline one"* in the sense that they are *"[t]echnically…a form of indirect identifier"* but the association between person and identifiers of this type may be strong, in particular because such identifiers *"will almost always be constructed out of more than one piece of information"* such that they have unique value functionally in many typical situations.

[744] Compare Booth et al (2004, p.87): "*[it] would seem that whether data that sometimes constitutes 'personal data' actually constitutes personal data in any given case depends upon the presence or absence of sufficient other information to enable the individual's identification. This of course raises the possibility that it is in fact this other information that is actually identifying the individual in the circumstances at hand and the data in question is simply linked to the individual through association with this identifying data".*

[745] Ibid, pp.108-109: *"things that operate as 'unique identifiers', in practice, are unlikely to consist of single pieces of data. They are rather more likely to constitute 'portfolios' of a number of pieces of data. These pieces of data may take the form of identifiers (which may, independently, be more or less unique in nature). It may be the more unique a specific identifier is then the fewer the number of additional identifiers it must be associated with, but, invariably, it will have to be joined with some additional information.* ***One of the questions that arise is whether information that is associated with identifying data may be 'personal data' even if the information itself is not capable of identifying the individual (in any context)"*** (emphasis added).

variables) in a relevant dataset context with relevant background knowledge.[746] Thus, "*it is not the uniqueness of the data per se, that is significant but the availability of a context (possibly informed by other identifiers) within which that data may function as a unique identifier*".[747]

(d) Jigsaw / mosaic effects identification

A related concept is jigsaw/mosaic-effects identification, the ability to identify someone from data (or re-identify them from data that has been subject to anonymisation techniques) using analytics on multiple data points, not possible considering the individual points by themselves.[748] Data linkability enables this by facilitating combinations that can create (i.e. logically associate with) new and non-superficial knowledge about an individual.[749] Whereas the link strength between an identifier and a person does not necessarily correlate with the sensitivity of what the identifier may reveal about someone alone.

2. Identity

A discussion about identifiers as things revelatory of people and reflecting aspects of data-person linkability raises the question, in turn, what might identification mean over and above this (especially in the context of considering the varying degrees of proximity that different types of data can have to particular persons)? First, however, it is useful to consider the concept of identity

---

[746] Sweeney (2001, p.21): *"combinations of characteristics can combine to construct a unique or near-unique identifier".* This is a particular concern when multiple anonymised datasets containing the same individuals are available to third parties, as they can be matched up (overlaid) with one another, to re-identify individuals. An example involved a study testing the re-identification possibilities of metadata datasets reported in De Montjoye et al (2015). Scientists in this study found that it is possible to determine the identity of shoppers using credit card purchase and location metadata. In particular, the study found that shopping receipts could be matched with four sources of external location data acquired from repositories like social media to determine identities with 90% accuracy. The so-called 'correlation attacks' permitting re-identification worked in part because each string of purchases was highly unique making identification possible with additional external metadata.

[747] Booth et al (2004, p.110) (in full): *"[e]ven the most 'common identifiers' might then be used as the 'key' to identification within an appropriate context…it is not the uniqueness of the data per se, that is significant but the availability of a context (possibly informed by other identifiers) within which that data may function as a unique identifier".*

[748] O'Hara (2011, p.11): "*[t]his is a technical concept describing the ability or otherwise of an adversary to reidentify or to deanonymise anonymised data, with the help of background information and processing power*".

[749] Unlinkability suggests the opposite. For example, see ISO/IS 15408: 1999: *"[unlinkability] ensures that a user may make multiple uses of resources or services without others being able to link these uses together. [...] Unlinkability requires that users and/or subjects are unable to determine whether the same user caused certain specific operations in the system*." Although to note, a distinction can be made between this absolute interpretation of unlinkability (no determination of a link between uses) and a relative interpretation of unlinkability (no change of knowledge about a link between uses). Compare also Hansen & Pfitzmann (2010), which defines unlinkability as an inability to distinguish relatability between a person and something. To note, the terms 'anonymity' and 'anonymisation' have been defined in terms of linkability. For example, respectively, as a "*condition in identification whereby an entity can be recognized as distinct, without sufficient identity information to establish a link to a known identity*" (ISO/IEC 24760-1:2011); and, as a "*process that removes the association between the identifying dataset and the data subject*" (ISO/TS 25237:2008). An interesting computer science analysis of this fact in the context of risk control over identity disclosures in a big data era is contained in Creese & Hodges (2013).

as anchoring the meaning of the verb 'to identify' (and thus 'identified' or 'identifiable'). This is because often interchangeable with the terms identification and identifiability is language around possibilities of using data to 'ascertain someone's identity'.

Identity is a multi-faceted notion, which is open to many interpretations in philosophy, sociology, and psychology literature, amongst others. Modern senses of the word can reflect social and cultural constructs, encompassing esoteric issues around who we truly are and think we are. However, dictionary definitions of identity also reflect earlier senses of the word in terms of the attribute of sameness. For example, in the Oxford English Dictionary 'identity' is, "*[t]he sameness of a person or thing at all times or in all circumstances; the condition or fact that a person or thing is itself and not something else; individuality, personality*."[750] Said otherwise, personal identity may be seen to relate to those set of attributes (qualities/properties) essential to a person being that person and no other in some non-contingent way.[751]

(a) Static notions of identity

Conceptions of identity, particularly in the latter sense, also often have a static component. At a basic and biological level, for example, our identity may be said to be unique to us and revealed by numerous objectively-verifiable biological features (e.g. our fingerprints or dental records). Conversely, we find 'identification' as a term used to denote connections made between information and a physical person interchanged with phrases such as, "establishing someone's real (or 'real-world') identity". However, knowledge of our body and our identity are not necessarily overlapping. In illustration, in an administrative context, a person's 'civil identity' is a term reminiscent of the (oft-legal) association of biographical identifiers (e.g. name, date of birth, address, and notational identifiers such as passport number) with particular persons to verify that they are who they say they are. The same term can also refer to the identity attributed to a person by a State (e.g., a natural person represented by a national insurance number, or a combination of details e.g. full name, date of birth, and birth location etc.). In both cases, "*an informational representation of the chain of life events*" (or at least some parts of that chain, as relevant to the context) may be adducible as in some way definitional and demonstrative of who someone is.[752]

[750] Oxford English Dictionary (1998)
[751] Compare Knight & Saxby (2014, p.618) where it is argued that Identity may also be conceived pragmatically as that "*which represents and makes us identifiable within a set of people*".
[752] Crosby (2008, p.9).

(b) Dynamic notions of identity

Our identity is often viewed dynamically and relatively (exhibiting a degree of fluidity liable to change over time) from a sociological/psychological perspective. For example, it could refer to a social category that we are in (or whose membership we feel we relate to), or cultural features we adopt. Indeed, we might be said to have multiple identities relating to the different roles we have in life (and the way we behave in each role), alongside other features such as our status and opinions.[753]

Suffice to say, however, an analysis of the noun 'identity' can only take us so far in examining its associated verb without it becoming tautologous (i.e. my identity is that by which I may be identified, and vice versa, data from which I can be identified is data from which my identity may be ascertained).

3. Identify

The verb 'to identify' may be considered to refer to the process of identification that can result in an output, that is, a state of being 'identified'. It is possible to conceive 'identifiability', in turn, as a state of being capable of being identified, and 'identifiable information' as that information by which an individual could be identified.

To delve deeper, there are (at least) three distinct but related meanings to the word 'identify', albeit with some lack of precision and clarity over how these meanings should be seen as interrelated. A description of each follows.

(a) Identify, as in to know accurately certain things about someone conferring equivalence

Identifying someone implies knowing something about an individual linked to him/her in some essential and accurate way, e.g. deducing something true about them based on certain information.[754] However, something further often seems necessary.[755] It implies an accurate

[753] As alluded to in Chapter 1, to note, this thesis is not concerned with identity in this sociological sense, or a possible third (esoteric and epistemic) way regarding the essence of who I am, e.g. from a Cartesian perspective or similar. Such meanings divert attention significantly away from the remit of this particular thesis that focuses on identification as a process that stand in relation to an object (data) and a subject (a particular person associated in some way with that data).

[754] Conversely, the concept of anonymisation is often conceived as suggesting the concealment of certain facts about someone – i.e. so that they can no longer be known (for further discussion of anonymisation concepts, see Chapter 5). Knowledge, in this sense, is not meant to preclude the possibility of machine-learning being encompassed within its ambit.

[755] As Zwenne (2013) points out "*[d]ata are not personal data just because they say something about an individual, even uniquely so, rather because they refer to someone whose identity is known or can be known*".

recognition of equivalence that – through this gained knowledge – it can be confirmed that someone is who they indeed are ('that is Alison'), and possibly further equated with someone else ('that is Alison, who is you').

Nonetheless, this meaning also begs some questions, such as regarding the point of equivalence at which knowing something (accurately) about someone from information becomes equivalent to deeming them identified from it, or at the very least capable of identification from it (identifiable).

Moreover, what must be the level of certainty or intentionality of thought process to cross the definitional line between being identified from data in contrast to only being identifiable from it, or is the difference one of objective fact (i.e. you are known when you are known, at which time you become identified, whereas previously you were unknown albeit knowable)? What exactly does knowing that someone is who they indeed are even mean, and require in terms of proof (particularly when based on multiple information pieces that singularly could not enable such knowing)? [756]

(b) Identify, as in to authenticate an individual as possessing certain attributes

A related interpretation of 'to identify' focuses on authenticating attributes belonging to someone against a certain standard.[757] Authentication here is a process that links claims to a person (e.g., do they have the right to engage in or perform some activity?). The conflation of the identification process with the authentication process can be seen from some definitions of authentication, such as being "an e*lectronic process that enables the electronic identification of a natural or legal person, or the origin and integrity of data in electronic form to be confirmed*".[758]

Yet, one might argue that authentication (of attributes) does not necessarily result in identification because, while authentication may involve verifying who we are (as us, and no other), it does not necessarily require this. For example, the matching of a biometric data sample with a biometric template already enrolled on a system may not – in itself - tell us anything more about the person

[756] In that sense, it seems clear that there is a question begged about whose knowledge is relevant and their pre-existing level of knowledge – that is, whether information that is associated with identifying data may be personal data even if the information itself is not capable of identifying the individual in any context. This is discussed further in Chapter 3.

[757] Knight & Saxby (2014, p.623). In a similar way, an 'identifier' has been defined as "*information used to claim an identity, before a potential corroboration by a corresponding authenticator*" (ISO/TS 25237:2008).

[758] Article 3(5), Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC (the EU 'eIDAS' Regulation). In Article 3 of that Regulation, 'electronic identification', in turn, is defined as meaning "*the process of using person identification data in electronic form uniquely representing either a natural or legal person, or a natural person representing a legal person*", whereas 'personal identification data' is described as "*a set of data enabling the identify or a natural or legal person…*".

who gave the sample,[759] similarly the presentation of a personal authentication token (such as entering a password, PIN code, or swiping a security pass).

However, some (typically more formal) authentication methods can rely on an individual knowing and recounting key personal details (such as their date of birth and address) and presenting documents corroborating their answers (e.g. a passport). Moreover, commonly equated are the act of recognition by the measurement/analysis of a biometric characteristic and a method of identification.[760] This is because of the near-unique nature of many physiological (and sometimes behavioural) underlying identifiers represented.

(c) Identify, as in to distinguish an individual from other individuals

This interpretation of 'to identify' focuses on the ability to distinguish an individual from other individuals in a dataset, or, in the world at large. In this sense, we might say someone is identified if they have been isolated from others in information, whereas they may be considered identifiable when there is the possibility to single someone out as unique within a group.[761] In minimum, there must be sufficient variability in the subjects to make distinction possible (i.e. between those members of the group of individuals about whom the information relates in the circumstances under consideration).

In this context, people's actions can also be a means to identification (doing not just being), as long as recognition from such behaviours is possible distinct from others. For example, individuals who interact online (e.g. in gaming communities) may say that they are able to identify each other in

[759] The House of Commons Science and Technology Committee (2015) has described biometrics as the *"science of establishing the identity of an individual based on the physical, chemical or behavioural attributes of the person".* Compare Article 29 Working Party (2012, WP193, pp. 3-4): "*biological properties, behavioural aspects, physiological characteristics, living traits or repeatable actions where those features and/or actions are both unique to that individual and measurable*"; although to note it also states (p.9): *"[h]uman tissue samples (like a blood sample) are themselves sources out of which biometric data are extracted, but they are not biometric data themselves (as for instance a pattern for fingerprints is biometric data, but the finger itself is not)".* Moreover, identity authentication via biometric data is only possible when there is a pre-existing sample to check against (i.e. a biometric sample must be placed within a context enabling the identification of an individual (see Knight & Saxby (2014, p.624)). It would depend upon a raw image being collected and enrolled in a system initially alongside identifying information, against which data (e.g., the digital code or template of a biometric feature, such as a fingerprint or iris or DNA) can be subsequently verified on later presentation.

[760] Compare Article 29 Working Party (2007, WP136, pp.5-6): *"[b]iometric systems are tightly linked to a person because they can use a certain unique property of an individual for identification and/or authentication. While a person's biometric data can be deleted or altered the source from which they have been extracted can in general neither be altered nor deleted. …biometric data, by their very nature, are directly linked to an individual".*

[761] Compare the following definition of 'identification' referred to as a *"process of using claimed or observed attributes of an entity to single out the entity among other entities in a set of identities*" (ISO/TS 25237:2008). More colloquially, the Merriam-Webster Dictionary [online] gives the following synonyms for the verb 'identify' (viz. "*to find out or establish the identity of*"): "*distinguish*", "*finger*" "*pinpoint*" and "*single (out*)".

their virtual environments – albeit possibly assisted in singling them out by online pseudonyms or other indirect identifiers - even if they would not be able to recognise them face-to-face.[762]

The problem with this meaning is that it applies even if the group is very small as long as one person has a unique identifier in a given context (i.e. sufficient to individuate a person from other group members). The likeness of this possibility of identification thus depends on the data set, its size, and the overlap of data between its entries. Moreover, some may argue that identification implies an element of real-life knowledge about someone over and above the capacity to single someone out.[763] For example, 'to identify' may be interpreted to refer to the ability to isolate someone in a group *and* contact them (e.g. by email, postal address or the URL of a social networking site profile).[764]

While Chapter 2 revisits some aspects of this discussion, the discussion in this Appendix admittedly concerns semantics that are not necessarily the same as legal meaning. However, the terminological confusion surrounding ambiguous terms – such as identification and identifiability – do not bode well for legal certainty in practice as discussed in Chapter 3.[765] Specifically, ambiguity seems to spring from the fact that the act of identifying someone often seems intrinsically related to who is carrying out the activity, the accuracy of his/her belief, and its evidential verification.

[762] Per Chapters 1 and 3, data gathered around online activity can be used to facilitate the virtual recognition of someone as the same person as another user via, e.g. a unique cookie value.

[763] Although, often, singling out online is done with the purpose of making a real-life factual connection. Thus, a possible concern for individuals related to the use of tracking software in respect of their online activities is that personal information may be inferred about them without their consent. For example, a person's online search queries could result in someone finding out who they are, even when different queries by the same person are recorded against a code number, where such information can be combined.

[764] Hamilton & Jay (2012, p.172): "*[a] person becomes identified where there is sufficient information either to contact him or to recognise him by picking him out in some way from others and know who he/she is*." See also, Dinant & Poullet (2004, p. 30), and Dinant (2010, p.7, translated into English): "*[t]his functional division of data actually distinguishes three types of personal data which are substantially different. They are, more precisely, properties of personal data. Thus, an email address such as "john.smith@coe.int" combines the three properties described above. We know that John Smith works at the Council of Europe. By typing his email address in a search engine, we can find related information and finally the email address allows us to contact John Smith, possibly for promotional purposes".*

[765] In addition, the potential ramifications of which meaning to attribute to the concept of identification are important ones. This is because different interpretations support different views regarding how far the application of data protection law extends (see Chapter 2).

Appendix 2 - The impact of Brexit on this thesis

The Brexit referendum vote (in June 2016) happened during the write-up stage of this thesis. This date was after research had already been carried out by the author into UK law in response to the research questions chosen, as an example of one EU MS's domestic data protection legislation in implementation of the DPD.

While the value in considering UK interpretations of the DPA is not undermined by the possibility of Brexit, with hindsight – in considering the value of this thesis prospectively – choosing a different EU MS as exemplar may have been preferable. For example, one possible future is where the courts of the UK will no longer have access to the CJEU and the CJEU will no longer be the final arbiter of the interpretation of data protection law in the UK.

Nonetheless, as the UK has not left the EU yet at the time of submission, and until it happens, the DPD~~A~~ continues to apply until that event. The GDPR will also automatically apply from 25 May 2018 as Brexit cannot have occurred by then as negotiations to leave the EU are expected to take some years.

Looking to the longer-term future, the impact of a Brexit is hard to predict at present and will very much depend on what is negotiated between the UK and EU MSs. However, data protection law is an area of law thought particularly unlikely to change post-Brexit in the UK.

For example, if the UK remains within the Single Market, the GDPR will continue to apply fully in the UK. In the alternative scenario, where there is a complete withdrawal, if the UK wishes to share data with EU MSs, or for it to handle EU citizens' data, they will need to be assessed as providing an adequate level of data protection (e.g. to a level equivalent to the high standards set out by the GDPR). In practice, this would likely require the UK government seeking an adequacy decision from the European Commission, declaring that the UK is "adequate" for data protection purposes.

Furthermore, whether the UK chooses to leave or to remain part of the EU, the GDPR rules may still apply to UK businesses. This is due to the fact that the GDPR has broad extra-territorial scope and will apply to organisations outside the EU who process the personal data of individuals in the EU, including where controllers offer goods or services to data subjects in the EU or where they monitor their behaviour, for example, through online profiling. In any event, many countries now have similar types of legislation modelled on the EU approach, and the trend is towards harmonising standards internationally in order to facilitate the safe flow of data across national boundaries.

Regarding the proposition to introduce a data protection block exemption (under Models 1 or 2) in Chapter 6, the effect of a hard Brexit would be that the UK would need to consider implementing a

national block exemption if it wished to emulate this proposition in national law (foregoing reliance upon the automatic application of a parallel exemption from UK data protection law resulting under an EU block exemption, the scenario as exists under UK competition law).

Appendix 3 – Additional UK case-law on the personal data concept under the DPA

Per Chapter 2, according to the DPA, personal data is data from which a living individual can be identified on its own or "*from those data and other information which is likely to come into the possession of, the data controller*". Thus, while section 1(1) DPA does not mention the term 'identifiable', it implies its relevance albeit only from the perspective of information which the data controller has access to, or may have access to, in the future. In other words, the primary focus of the definition is on the controller and that information which is (or may come into) its possession. This is typically referred to as the 'in the hands of [the data controller]' concept, a phrase made famous by Lord Justice Hope in a 2008 case described below, later referred to in two significant cases dealing with similar issues also described.

Below is a brief description of three notable cases including UK judicial interpretation of the personal data concept under the DPA. All relate to factual scenarios involving the potential for re-identification risk from statistically-anonymised data disclosed to third parties. To complement the analysis carried out in Chapter 3, considering the interpretations given in these cases may also help decipher the correct approach to determining the issues of 'identifiable how', 'identifiable from whose perspective', and 'identifiability with what likelihood of occurring' under UK law.

1. Common Services Agency v Scottish Information Commissioner

This relevant judgement was issued by the House of Lords (now Supreme Court): Common Services Agency (Appellants) v Scottish Information Commissioner (Respondent) (Scotland) [2008] UKHL 47. In background, the facts involved a request made to the Common Services Agency (CSA) under the Freedom of Information (Scotland) Act 2002 (FOISA).[766] The request related to the incidence of childhood leukaemia in a particular postal area connected with statistical information gathered from Scottish health boards. The CSA refused to share the statistical information on the basis that there was a risk of re-identification of living individuals (particular child leukaemia sufferers) from

[766] Under section 38 of FOISA – like section 40 FOIA - the test for deciding whether personal data can be disclosed is whether disclosure to a member of the public would breach the data protection principles. In other words, both legislative provisions are similarly intended to ensure that public authorities take into account the additional information that a particular member of the public might have that could allow data to be combined to produce information that relates to and identifies a particular individual – and that is therefore personal data.

that data.[767] For that reason, the CSA deemed the information personal data under the DPA, and accordingly disclosure-exempt under section 38 of FOISA.

The Scottish Information Commissioner subsequently ordered the CSA to perturb the figures using a process called barnardisation – a method of disclosure control for tables of counts that involves randomly adding or subtracting one from some cells in the table - which would hide the precise figures but reveal the general pattern of leukaemia. The decision containing this order was appealed by the CSA, first to the Inner House of the Court of Session (which upheld it),[768] and then to the UK House of Lords (which overruled it).[769]

The key issue in this case was that - putting aside the fact that it was deemed likely that the modified data would not be personal data in the recipients' hands - the same was not (necessarily) true for the CSA. This was because it retained the raw data underlying the modified information (which might allow it to identify particular children from this modified information). In other words, the case was concerned with the modified information's status and, in particular, whether an objectivist or relativist perspective on identifiability in respect of such information was the correct approach to take, under UK law.

Several different approaches were suggested by the Lordships in holding that the modified information would *not* be personal data when publicly disclosed (even though the DPA's definition of personal data (paragraph (b)) requires account to be taken of other information which is in, or is likely to come into, the possession of the data controller). Thus, the Lordships appear to be approving a relativist approach to the issue of identifiability. For example:

- Lord Justice (LJ) Hope of Craighead commented that it should be considered whether the other information held by the CSA (that is, the raw data) would add anything to the modified statistical information that – in combination - would enable the identification of individuals. If it would not, the statistics would not be personal data.[770] On the other hand,

[767] For example, pointing to the low numbers of individuals involved in the raw data, as well as the inclusion of information on rare diagnoses, that may provide outlier points in the information that might enable individuals to be singled out from the data by the recipient post-disclosure.
[768] Common Services Agency v Scottish Information Commissioner IHCS [2006] ScotCS CSIH 58
[769] The House of Lords (Supreme Court) then remitted the case back to the Scottish Information Commissioner to decide as a question of fact whether the information in a barnardised form was personal data if read together with other information also held by the CSA.
[770] Common Services Agency (Appellants) v Scottish Information Commissioner (Respondent) (Scotland) [2008] UKHL 47, para. 27: *"[i]n this case it is not disputed that the Agency itself holds the key to identifying the children that the barnardised information would relate to, as it holds or has access to all the statistical information about the incidence of the disease in the Health Board's area from which the barnardised information would be derived. But in my opinion the fact that the Agency has access to this information does not disable it from processing it in such a way, consistently with recital 26 of the Directive, that it becomes data from which a living individual can no longer be identified. If barnardisation can achieve this, the way will be then open for the information to be released in that form because it will no longer be personal data".*

"*if the "other information" is incapable of adding anything and "those data" by themselves cannot lead to identification, the definition will not be satisfied" because such "other information" would then have had "no part to play in the identification*".[771] In other words, the raw data must play an *operative part – in combination with the modified data* - in enabling a person to make an identification (such as in circumstances where the statistical information can still be cross-referenced against the raw data held by the data controller). LJ Hope endorses a relativist approach, as evidenced by the following comment: "*[i]f it was impossible for the recipient of the barnardised data to identify those individuals, the information would not constitute "personal data" in his hands*".[772]

- Baroness Hale (now President of the UK Supreme Court) adopted a different view, but with a similar relativistic approach. She said that, whereas an authority disclosing modified data might have the 'key' to link the data back to individuals and remained bound by the data protection principles, this is not true for the recipient of such data from the authority. In other words, "*the recipient of the information will not be able to identify the individuals either from the data themselves, or from the data plus any other information held by the Agency, because the recipient will not have access to that other information*".[773] Hence, as the recipient would not be able to identify individuals from it, for the purpose of disclosure, she held that personal data was not being processed.

2. Department of Health v Information Commissioner and another

A similar factual scenario was considered in a case before the English High Court in Department of Health, R (on the application of) v. Information Commissioner [2011 EWHC 1430 (Admin) (20 April 2011). The facts of this case were that, after the UK Department of Health (DOH) changed their approach to the release of anonymised abortion statistics, the Pro Life Alliance requested from the DOH, under freedom of information rules, anonymised statistics in the more detailed format in which they had previously been released. The DOH refused that request on the basis that the DPA

[771] Ibid, para 24.
[772] Ibid, para 26.
[773] Ibid, para 92 (in full): "*I am assuming the particular data which Mr Collie has requested, anonymised in such a way that neither he nor anyone else to whom he might pass them on could identify the individuals to whom they relate. The Agency may well have the key which links those data back to the individual patients. The Agency therefore could identify them and remains bound by the data protection principles when processing the data internally.* ***But the recipient of the information will not be able to identify the individuals either from the data themselves, or from the data plus any other information held by the Agency, because the recipient will not have access to that other information***. ***For the purpose of this particular act of processing, therefore, which is disclosure of these data in this form to these people, no living individual to whom they relate is identifiable"*** (emphasis added).

prevents the disclosure of personal data under FOIA 2000 where this would breach data protection principles.

In 2006, the Information Commissioner (IC) investigated the reasons given by the DOH for its refusal. The IC issued its decision 2008, concluding that the disputed information was not personal data, as it was not possible to identify either doctors or patients from the statistics. On appeal, the Information Tribunal disagreed, finding that the statistics constituted personal data (although it also concluded that there was insufficient risk of re-identification).

On a second appeal, the High Court upheld the original IC decision, holding that the requested information was not personal data and could be disclosed. In other words, even though the DOH had additional information – which if disclosed would enable the identification of the individuals who had certain types of abortion – the Court concluded that the statistics alone could not amount to personal data. Again, a relativistic approach is being used here. The court held that recipients of anonymised abortion statistics could not identify individuals, and therefore that the anonymised data was not personal data for the purposes of the DPA and the section 40(2) exemption. The High Court judge, Cranston J, also stated that, to consider the requested data as personal data, would establish a principle that would prevent any publication of medical statistics, however broad. Also stated was the opinion that the risk of identification must be greater than remote and reasonably likely for information to be classed as personal data under the DPA.[774]

3. All Party Parliamentary Group on Extraordinary Rendition v The Information Commissioner & The Ministry of Defence

A third case of interest related to this interpretation issue is an Upper Tribunal decision from 2011: All Party Parliamentary Group on Extraordinary Rendition v. The Information Commissioner & the Ministry of Defence [2011] UKUT 153 (ACC). The All Party Parliamentary Group on Extraordinary Rendition (APG) - a cross party group of MPs and peers - requested various information relating to the treatment of persons detained in the conflicts in Iraq and Afghanistan from the Ministry of Defence (MOD) under FOIA 2010. The Upper Tribunal held that the disclosure of data in statistical form pursuant to this request did not constitute processing of personal data, so that the disclosure did not have to comply with data protection principles.

[774] Para 16. To paraphrase, the argument was put forward that statistical data (including information as to different foetal abnormalities and the total number of terminations) was withheld on the grounds that detailed statistics could, given the small numbers involved in some categories, risk identification of patients or doctors. This argument was rejected on the evidence by the Court - it decided that the risk of identification was extremely remote, therefore Article 8 was not engaged and disclosure was ordered.

The Tribunal said:

> [O]utside the hands of the data controller the information is no longer personal data, because no individual can be identified … the best analysis is that disclosure of fully anonymised information is not a breach of the protection of the Act because at the moment of disclosure the information loses its character as personal data. It remains personal data in the hands of the data controller, because the controller holds the key, but it is not personal data in the hands of the recipients, because the public cannot identify any individual from it. That which escapes from the data controller to the outside world is only plain vanilla data.[775]

4. Critical summary

The above case-law suggests that a hard-line judicial approach (whereby statistical data would remain personal data) has been rejected by the English courts. Thus, in the UK, a data controller should be able to aggregate originally-personal data and then disclose or process them as statistics because the public cannot identify living individuals from such data. Yet, while the controller still holds the original data, it must process that raw data as personal data in compliance with the DPA. In other words, this analysis confirm the conclusion taken in Chapter 3 that the personal element of data that has been anonymised in statistical form is considered from a relativistic (subjective) viewpoint on identifiability under UK law generally. A controller may release such data without breaching the DPA because, on disclosure, the data would, in the hands of third parties, be deemed to lose its personal data status.

Notwithstanding, as pointed out in Chapter 3, statistical (or irreversibly encrypted) data is different in nature from – and arguably poses fewer risks than – data where direct identifiers have been removed but individual-level indirect identifiers remain. Arguably, there are stronger arguments that can be put forward that statistical or irreversibly encrypted data are no longer personal data than is the case with data that retains indirect identifiers where the risks involved are higher (e.g. as third parties can single out individuals from the data and potentially re-identify them using information outside the control of the data controller). There is an additional UK-decided case not so far mentioned dealing with the latter scenario, however, as it is an old case predating the emergence of big data analytics and the ratio decidendi of the judgement lacks precision (focusing on the facts instead), it is mentioned next in this Appendix for completeness only.

[775] Paras 127-128.

5. Department of Health, ex parte Source Informatics Ltd

In this English case (considered in the High Court and the Court of Appeal in 1999), it was argued that removing the name of a patient from a health prescription was sufficient to take its future processing outside the remit of the DPA. In the High Court (Department of Health Ex p. Source Informatics Ltd (No.1), R. v [1999] EWHC 510 (Admin) (28 May 1999), Latham, J. said:

> So far, I have discussed the case on the basis that anonymity can be guaranteed. However, the applicants have themselves accepted that there is a remote risk that certain information of a rare kind might conceivably enable a patient to be identified. I fully accept that there is no evidence before me which sets out any rational basis for such concerns. Nonetheless, it highlights the fact that systems may not always be perfect. In these circumstances, why should the patient be deprived of the opportunity of making up his own mind as to the risk, such as it may be? This approach also has the merit, it seems to me, of placing the debate in its correct context.

Said otherwise, the judge noted recognition of *"a remote risk that certain information of a rare kind might conceivably enable a patient to be identified*" – however, he accepted that there was "*no evidence before me which sets out any rational basis for such concerns*" so he decided the patients' anonymity *could* be guaranteed.

On appeal, Brown LJ in the Court of Appeal also reached this conclusion (in Source Informatics Ltd, Re: An Application for Judicial Review [1999] EWCA Civ 3011 (21 December 1999)). However, it was also noted (para.7) that it would be for the applicant Source Informatics "*to satisfy all interested parties that there will be no risk of identification in practice. Source are confident of achieving this and have recently proposed certain refinements to their system to screen out any conceivably identifying information*".

Appendix 4 – The Effects-based Approach and the non-personal data concept

This Appendix considers the advantages and disadvantages of using the Effects-based Approach as developed in Chapter 4 for determining when to deem data sufficiently non-personal to fall outside of data protection rules. Assessment is made against the standard of how best it might achieve data protection's twin objectives in that context (while maintaining broad coherency with existing data protection terminology/principles). While the Effects-based Approach is rejected at the end of Chapter 4 onwards, the following analysis is useful to the reader as a summary of how regulators and others have linked the concepts of anonymisation and negative effects in the past. The analysis also highlights how an (re-)identificatory analysis and a harms-based one are inextricably linked in case law, regulatory thought, and in theory.

To develop the argument that the Effects-based Approach is better than the identificatory-approach, under consideration in this Appendix is the proposition that by focusing directly on the mitigation of harm resulting from particular personal data processing activity, the Effects-based Approach is able to sidestep formalistic problems around interpreting identifiability. Also explored is the proposition that a concept of non-personal data based on usage-effects fits better with the realities of dynamic (rather than static) data contexts, reflecting the notion of data anonymisation taking place in data 'environments'[776] for consideration holistically in the relevant circumstances.

To recap, under the Effects-based Approach, to determine if there is a Relevant Effect denoting that the processing of data relating to persons should be deemed personal data, data controllers would be required to assess, not just the risk (likelihood) of harm occurring from a particular data-usage, but also the likely severity of that harm.

Conversely, to determine if non-personal data exists following the application of anonymisation techniques, requires consideration of what it means for a particular processing activity planned in respect of data relating to a person to *no longer* be deemed liable to have a Relevant Effect upon that person (whereas previously – pre-modification – it was so deemed liable). In this context, it is possible to talk about data controllers taking measures to mitigate both: the magnitude/severity of harm likely to result so that it would no longer be assessed as appreciable; and, the chance of harm

[776] Elliot et al (2016, p.2 and p.68) adopt the term 'data environment', which they use in reference to a totality of "*people, other data, infrastructure, and governance structures*" for assessment in respect of data to ascertain realistic measures of risk (albeit they are referring solely to post-anonymisation data disclosure risk). In other words, it is only by considering relevant data and its environment in totality (what they call the 'data situation', p.xiii) that the need for - and extent of - anonymisation efforts and its efficacy can be assessed.

occurring to the point where it is no longer reasonably likely to flow from the data processing activity. Both activities are by way of a kind of ex-ante negative effects risk-management.

A. Support for the Effects-based Approach regarding the non-personal data concept under the DPD/DPA

We turn to authoritative support for the Effects-based Approach under EU/UK data protection law in the context of determining the status of data that has been subject to anonymisation techniques. Per Chapter 4, risk-based discourse is embedded in many ways in current law and the GDPR, but to what extent can this be said to be a risk-of-harm (negative effects) discourse implicitly, rather than a re-identification risk issue, when it comes to the concept of non-personal data?

The WP

It is worth revisiting some risk-based remarks in previously-discussed WP opinions in the context of examining them vis-a-vis the Effects-based Approach and potential support for it implied within them.

WP 136 (2007)

WP136 refers to 'back-tracking-to-identity' risks from a type of indirectly identifiable information called 'retraceably pseudonymised data', the processing of which the WP states implies "*risks at stake for the individuals*" (albeit most often low ones).[777] Said otherwise, the WP suggests, not only that different levels of identifiability can be associated with data that has been subject to anonymisation processes, but also - depending on the strength of the link between such information and the individual to whom it relates – its processing can have different levels of risks of harm to such individuals. Again, however, the nature of the "*risks at stake for individuals*" being alluded to by the WP is not entirely clear.[778]

---

[777] WP136, p.18: *"[r]etraceably pseudonymised data may be considered as information on individuals which are indirectly identifiable. Indeed, using a pseudonym means that it is possible to backtrack to the individual, so that the individual's identity can be discovered, but then only under predefined circumstances. In that case, although data protection rules apply, the risks at stake for the individuals with regard to the processing of such indirectly identifiable information will most often be low…*" The quote ends as follows: *"so that the application of these rules will justifiably be more flexible than if information on directly identifiable individuals were processed."* See further discussion in Chapter 7, alongside a quote at p.19 of WP136: "*[a] different question is that those data protection rules could take into account whether risks for the individuals are reduced, and make processing subject to more or less strict conditions, based on the flexibility allowed by the rules of the Directive*". In comparison, regarding anonymisation, the ICO has pointed out different levels of identifiability and argued for a more nuanced and contextual approach to the protection of personal data and anonymised data.

[778] Compare ibid, p.25: "*[t]hese rules were therefore designed to apply to situations where the rights of individuals could be at risk and hence in need of protection".* By comparison, the WP said in 2015 (Article 29 Working Party. Annex to Letter to Ms Ilze JUHANSONE Ambassador Extraordinary and Plenipotentiary Permanent Representative to the EU,

WP216 (2014)

In WP216, the WP recognises explicitly the negative effects that may befall individuals if anonymisation is done incorrectly, citing the "*many examples of incomplete anonymising entailing subsequent adverse, sometimes irreparable effects on data subjects*".[779] Furthermore, the WP states even "*properly anonymised data (especially in the case of profiling)*" requires consideration of impact on individuals (i.e., "*even though data protection laws may no longer apply to this type of data, the use made of datasets anonymised and released for use by third parties may give rise to a loss of privacy)*". [780] For this reason, the WP suggests *"[s]pecial caution is required in handling anonymised information especially whenever such information is used (often in combination with other data) for taking decisions that produce effects (albeit indirectly) on individuals*".[781] Simply put, according to the WP, data controllers should undertake a negative-effects risk-assessment before they apply anonymisation techniques to any personal data they hold. This is not least because the processing of modified data (post-anonymisation, whether carried out effectively or not) can still present residual risks of harm to data subjects.

What might these harms be? The WP comments:

> [B]eyond the direct impact on data subjects produced by the consequences of a poor anonymisation process (annoyance, time consumption and feeling of lost control by being included in a cluster without awareness or prior consent), other indirect side effects of poor anonymisation may occur whenever a data subject is included in a target erroneously by some attacker, as a consequence of processing anonymised data - especially if the attacker's intents are malicious. [782]

Furthermore, the WP links the switching on/off of the application of data protection rules to the ways that modified data are used. It states, "*various use cases can be envisaged for anonymised data, ranging from social surveys, statistical analyses, new service/product developments*" and *"[s]ometimes, even such general purpose* ***activities may have an impact***

Brussels, 17 June 2015, pp.5-6): "*pseudonymising techniques used to disguise identifies…can help to reduce risks to individuals*" and, consequently, it encourages the application of pseudonymisation techniques as a "*privacy tool [that] helps to minimize the processed information and, subsequently, the risks (e.g. in the scientific sector where 'key coded data' is processed*". See also p.13 of the same document: "*the applicable safeguards taking into account the nature, scope and purposes of the processing and risks for the rights and freedoms of the data subjects"*; and pp.15-16: "*the data protection risks represented by the processing and the nature of the personal data".*

[779] WP216, p.9. In that context, the WP states that data controllers should balance their anonymisation effort and costs (in terms of both time and resources required) against these many examples, as well as the increasing low-cost availability of technical means to identify individuals in datasets, and the increasing public availability of other datasets.

[780] Ibid, p.11.

[781] Ibid, p.11.

[782] Ibid, p.23. The WP also comments (p.3) that "*anonymisation techniques can provide privacy guarantees, but only if their application is engineered appropriately*…"

on specific data subjects, nullifying the supposedly anonymous nature of the processing data" (emphasis added).[783]

The ICO

The ICO has linked more explicitly the concept of non-personal data and the risk of negative effects upon individuals flowing from processing activities. In 2014, it made a direct link between non-identifiability and an absence of direct effect upon individuals from information relating to them (albeit that it omits reference to how processing plays a part in the causation of effect):

> [O]nce information ceases to identify anyone **then it ceases to have any direct effect on them**, **so it poses a lower privacy risk**, meaning researchers and others should be much freer to use it for their own purposes than information held in a personally identifiable form.[784] (emphasis added)

Statements made elsewhere by the ICO, as considered next, may explain its rationale for that statement.

Anonymisation code of practice (the Code, 2012)

In the Code, the ICO addresses head-on the issue of how to determine whether someone is re-identifiable from post-anonymisation modified personal data under the DPA. In this context, it indicates additional (non-DPA-permissible) factors considered relevant to its own case determinations. These include the assessment of effects for the individuals to whom data for processing relates consequential upon their identification (so-called 'secondary harm effects').[785] Briefly, the ICO believes it good practice, when releasing anonymised data, to try to assess "*what*

---

[783] Ibid, p.27 (in Annex 'A Primer on Anonymisation Techniques).
[784] ICO (2014, SMD0018).
[785] The Code, p.20: *"[t]he test in the DPA for determining whether information relating to a living individual is personal data is based entirely on the identification or likely identification of the individual. The risk posed to individuals by disclosure, or the public benefit of this, are not factors that the DPA allows to be taken into account when determining whether or not information is personal data. In reality though, some types of data will be more attractive to a motivated intruder than others and more consequential for individuals. In reality these factors should also inform an organisation's approach to disclosure.* ***Clearly the identification of an individual can have a range of consequences depending on the nature of the data, the context in which it is disclosed and who it is about. The Information Commissioner would certainly be more concerned about a disclosure of personal data that is detrimental to an individual, than about an inconsequential one. The Information Commissioner will take the effect or potential effect into account should a case of reidentification or inappropriate data disclosure come to his attention. In borderline cases where the consequences of re-identification could be significant eg because they would leave an individual open to damage, distress or financial loss, organisations should: seek data subject consent for the disclosure of the data, explaining its possible consequences; adopt a more rigorous form of risk analysis and anonymisation. In some scenarios, data should only be disclosed within a properly constituted closed community and with specific safeguards in place. In some particularly high-risk situations, it may not even be possible to share within a closed community****"* (emphasis added).

the consequences of re-identification are likely to be, if any, for the data subject concerned",[786] following this approach itself in determining whether to take enforcement action related to instances of inadequate anonymisation.[787] Notwithstanding, the ICO adds that assessing likely consequences, "*can be difficult to assess in practice and a member of the public's sensitivity may be different from yours. For example, the disclosure of the address of a person on a witness protection scheme could be far more consequential than would usually be the case*."[788]

The Code also refers to effects in the context of discussing the motivated intruder test. It suggests that data 'impactfulness' is the most likely attractive factor potentially motivating an attempt at de-identification. Other factors include the possibility of causing harm as a desired outcome (including, presumably, harm to the individual who relates to the data under consideration).[789] This approach suggests the ICO is trying to 'square a circle' here by aligning the existing identificatory-approach to personal data with an impact-aware regulatory model (something akin to the Effects-based Approach in essence).

Yet, the exact relationship between the two types of risk remains elusive in the Code. Sometimes, the ICO seems to equate re-identification risk with "*privacy risk*" (whatever that might be intended to mean).[790] Other times it describes them as conceptually distinct, yet simultaneously confirms that making a positive assessment that personal data exists based on the level of re-identification risk from data in practice requires assessing possible disclosure-impact first:

> The possibility of making an educated guess about an individual's identity **may present a privacy risk but not a data protection one because no personal data has been disclosed to the guesser**. Even where a guess based on anonymised data turns out to be correct,

---

[786] The Code, p.25.

[787] Compare the Code, p.27: *"[t]he Information Commissioner is confident that adopting the techniques and procedures recommended in this code will guard against re-identification. However, in some cases re-identification may be a possibility.* ***Where there is evidence of re-identification taking place, with a risk of harm to individuals, the Information Commissioner will be likely to take regulatory action****, including the imposition of a civil monetary penalty of up to £500,000."* (emphasis added)

[788] The Code, p.25.

[789] The Code, p.23: "*[c]learly, some sorts of data will be more attractive to a 'motivated intruder' than others. Obvious sources of attraction to an intruder might include: finding out personal data about someone else, for nefarious personal reasons or financial gain;* ***the possibility of causing mischief by embarrassing others; revealing newsworthy information about public figures; political or activistic purposes, eg as part of a campaign against a particular organisation or person;*** *or curiosity, eg a local person's desire to find out who has been involved in an incident shown on a crime map….* ***data with the potential to have a high impact on an individual is most likely to attract a 'motivated intruder'***" (emphasis added).

[790] See also, e.g., the following quote from the Code (p.25) that seems to equate re-identification risk with 'privacy risk' (without it being clear precisely what this term means): "*[t]he risk of re-identification posed by making anonymised data available to those with particular personal knowledge cannot be ruled out, particularly where someone might learn something sensitive about another individual if only by having an existing suspicion confirmed. However,* ***the privacy risk posed*** *could, in reality, be low where one individual would already require access to so much information about the other individual for re-identification to take place*" (emphasis added).

> this does not mean that a disclosure of personal data has taken place. **However, the consequences of releasing the anonymised data may be such that a cautious approach should be adopted, even where the disclosure would not amount to a disclosure of personal data.** Therefore it may be necessary to consider whether the data should be withheld for some other reason, as discussed later in this code. This is clearly a difficult area of the law and **in approaching questions of disclosure it can be helpful to look primarily at the possible impact on individuals and then to move on to the more technical issue of whether or not there is likely to be a disclosure of personal data subject to the DPA."** [791] (emphasis added)

Big data and data protection (2014, rev.2017)

In 2014, the ICO associated the application of anonymisation techniques to, not just mitigating re-identification risk, but also mitigating at the same time data misuse risks:

> Anonymisation should not be seen merely as a means of reducing a regulatory burden by taking the processing outside the DPA. **It is a means of mitigating the risk of inadvertent disclosure or loss of personal data**, and so is a tool that assists big data analytics and helps the organisation to carry on its research or develop its products and services. [792] (emphasis added)

The ICO also describes anonymisation techniques as a type of privacy-by-design solution "*intended to protect privacy by mitigating the risk of re-identification and the risk of data misuse*".[793] The ICO clearly associates matters of jurisdictional assessment (i.e. when data may be considered no longer subject to the DPA) and those of substantive assessment (i.e. what needs to be done to ensure that data protection rules are complied with if they were so deemed to apply).

---

[791] The Code, p.26. Compare, p.7: "*[w]e draw a distinction between anonymisation techniques used to produce aggregated information, for example, and those – such as pseudonymisation – that produce anonymised data but on an individual-level basis.* ***The latter can present a greater privacy risk, but not necessarily an insurmountable one"*** (emphasis added). See also p.26: "*[i]nformation about groups of people - In some circumstances the release of anonymised data* ***can present a privacy risk even if it does not constitute personal data and cannot be converted back into personal data****. This might be the case where the anonymised data points to a number of individuals, eg the occupants of a group of households or those living within a particular postcode area. Information that enables a group of people to be identified, but not any particular individual within the group is not personal data. Conversely, information that does enable particular individuals within a group or all the members of a group to be identified will be personal data in respect of all the individuals who can be identified".*

[792] ICO (2014, #1, para.46, p.13). In the 2017 version of this document, this statement is copied with minor amendments (ICO, 2017, #2, para. 139, p.61).

[793] ICO (2014, #1, para.103, p.32): "*[p]rivacy by design solutions can involve not only* ***anonymisation techniques*** *but a range of both technical and organisational measures. These include access controls and audit logs, data minimisation, data segregation and purpose limitation and separation. These are* ***intended to protect privacy by mitigating*** *the risk of re-identification and* ***the risk of data misuse***" (emphasis added). In the 2017 version of this document, however, the last sentence is deleted and the rest of the quote is amended (ICO, 2017, #2, para. 165, p.73).

Judicial guidance

It is worth revisiting the FTT judgement in the *QMUL* case, where we can find hints that effects-related considerations are important when determining whether data are personal or not.

QMUL raised concerns that any FOIA disclosure of the modified trial-participant data would be without limitations as "*there would be no controls on ethical use, access, security, continue obligations of confidence or its further processing of any kind*" (and it could easily end up becoming public, e.g. through being placed on a website). [794] Consequently, QMUL argued that there was a risk of participant re-identification, but also risks of primary and secondary harms that might befall them post-disclosure.[795] The FTT responded, "*this is all predicated on the inability to sufficiently anonymize the data"*,[796] while ultimately holding that it was satisfied that the data had "*been anonymised to the extent that the risk of identification is remote*".[797] Said otherwise, in dismissing the appeal, the FTT appears to take position that, because re-identification risk is remote on the case facts, then at least in part this means that any risk of participant harm is also remote.

However, the FTT also went further. It commented on the likelihood of secondary harm befalling the participants when it noted QMUL's acceptance of the fact that "*there is no evidence of any threats by activists of physical violence against participants*" (albeit that QMUL had argued that, on the facts, it was reasonable to expect that some campaigners would be strongly motivated to re-identify them).[798] According to the FTT, if there is no evidence of hostility towards participants on the facts, moreover, then the risk that activist groups will recruit a sympathetic NHS-insider to access relevant records is far-from-obvious.

How the Effects-based Approach would apply in resolving whether the requested data was personal data is an interesting thought-experiment and way to explore how it would apply to secondary

---

[794] IC decision FS50565190, p.16.
[795] FTT judgement, p. 33.
[796] Ibid, p.33: *"(1) Should anonymization fail, conclusions may be drawn about individuals. (2) An individual could be identified as one of a group who may well have been participants, short of definitive identification, which is adverse in terms of individual privacy even if that information is not personal. (3) Individuals suffering anxiety and distress for fear that they might be identified and/or disclosure may be contrary to their expectations. (iv) individuals are exposed to a risk of future identification*". To note some primary, not secondary (conditional on identification taking place) harm was also noted, e.g. regarding the possibility for publicity-surrounding disclosure causing anxiety to participants (although linked to the fact that potential hostility could be shown to them by certain groups). Compare IC decision FS50565190, p.24: *"[i]n the other case, the University informed the Commissioner that the withdrawal was linked to concerns about confidential data being used by researchers and other people, even in an anonymised form. The University contended that, although it could not make a direct link to a data release in this case, it did not take much imagination to conclude that news of any PACE data being released, following an FOIA request, would have had a detrimental effect on this ex-participant, and may have led to, or encouraged, them to request that all their data be destroyed".* See also the statement by QMUL in IC decision FS50565190 at p.16: "*a breach of the patients' expected reasonable expectations of confidentiality in light of their consents to participate in PACE in reliance on non-disclosure except to limited practitioners and researchers"*.
[797] Ibid, p.39.
[798] Ibid, p.33.

harm considerations. Specifically, where the processing activity under consideration would be the act of (unrestricted) disclosure of the modified participant-related data to the FOIA requester, under the Effects-based Approach would this data be deemed personal data? If not, why not?

A point of agreement in the case was that the data requested was sensitive personal data (being health-related). Per Chapter 4, under the Effects-based Approach, it is proposed to include a working presumption that where processing involves sensitive data types, personal data exists. This is to address the risk of false negatives, as the harm magnitude likely to flow from sensitive data processing is often adjudged appreciable on the facts (even if the probability of the harm occurring is deemed not reasonably likely on the facts). Said otherwise, the data requested applied to the case facts would be deemed personal data.

However, adding an extra limb to the Effects-based Approach could be useful here. That presumption could be made rebuttable on the evidence, including in respect of situations specifically where secondary harm (harm consequent upon identification occurring) is involved. Moreover, to this end, use of a motivated intruder 'style' test could be encouraged where the substance of assessment would be changed (from assessing re-identification risk) to considering the likelihood of - including motivation and means through which - a hypothetical third party might cause secondary harm to the data subject if re-identification were to take place.[799]

By way of example, if the data subjects were known vivisectionists, releasing data to an activitist (or, indeed anyone, motivated to put it into the public domain) could result in appreciable harm befalling them precisely because of evidenced hostility towards this category of persons (e.g. by animal rights activitists). By contrast, in another situation (such as was found on the facts by the FTT in the *QMUL* case) sufficient evidence of hostility towards the individuals concerned - suggesting that appreciable harm was likely to befall them following the data disclosure - may *not* be found. Said otherwise, such substantiated evidence could be deemed sufficient to rebut the presumption that personal data exists (because a category of sensitive data was involved), precisely because disclosing the relevant data is not likely reasonably to give rise to appreciable harm (effects upon) the individuals.

[799] For example, relevant considerations might include questioning whether the data relating to persons to be processed and the processing activity intended suggest that: individual consumer profiles could be inferred for price discrimination; it involves financial or other information enabling identity theft; it involves information that could be used to blackmail individuals or to discriminate against them; it involves medical information that could be used by insurance companies (e.g. to deny coverage based on a pre-existing medical condition); or, inferences about creditworthiness could be used to assess credit risks and take negative decisions about an individual; etc.

Furthermore, assuming that the relevant data is in fact non-sensitive (in a legal definitional sense)[800] - removing the legal presumption that personal data is involved – a decision-maker such as the FTT would instead have to revert to the two-stage test under the Effects-based Approach:[801]

1) whether the probability of harm occurring (flowing from the relevant data's disclosure) is reasonably likely, but also:

2) whether the magnitude of that harm would be appreciable.

Said otherwise, secondary harm assessments would necessarily entail consideration of the likeliness of identification actually happening related to limb (1). To use the above example, if the data subjects were known-vivisectionists, before disclosing modified data to a FOIA requester, it would be necessary to consider the likeliness of re-identification taking place. If, on the facts, it were considered sufficiently likely to occur - and appreciable harm were deemed likely to befall the data subjects because of assumed hostility – then, the data would be personal data. On the other hand, that would not be the case if there were not sufficient evidence of hostility towards the data subjects such that appreciable secondary harm would be unlikely to befall them following the data disclosure, *even if re-identification were sufficiently likely*. Applied to the facts of the QMUL case and the FTT's level-of-risk analysis preliminary conclusions (and still assuming that the requested data had in fact not fallen within the sensitive data category), therefore, its conclusion that personal data were not involved would have been justified under the Effects-based Approach.[802]

However, it is clear that this proposed decision-tree framework cannot sidestep the interpretative challenges of the identificatory-approach altogether when it comes to secondary harm. Assessing re-identification risk associated with data disclosure (limb (1)) is a constituent part of the Effects-based Approach in this context (notwithstanding, the underlying focus is on mitigating harm

---

[800] Which in reality is also arguable on the facts if you look at it from an impact point of view, taking into account that there was no sensitive knowledge that could be gained from the trial data disclosed in and of itself. Contrast if the data output had been HIV status. On the other hand, some may argue that the fact someone has ME could be deemed very sensitive in terms of the effect it can have e.g. on one's recruitment prospects.

[801] This assumes that the reasonable likelihood of appreciable primary (non-identification dependent) harm befalling the data subjects following the processing activity of disclosure had already been assessed and discounted on the facts. Thus, only secondary harm is being considered here.

[802] This is disregarding the argument that could be made that participants may have been imbued with reasonable expectations that data relating to them would not be shared because of the confidentiality agreement signed by QMUL (and *may therefore* have been considered likely to suffer some element of emotional harm when that agreement was breached, and potentially – although not obviously - appreciable harm). Notwithstanding, this is not discounted as relevant factor that might be considered in some cases. Compare Elliot et al (2016, p.2): "*[t]hinking about the impact side of risk brings us to the third key concept, sensitivity, which tends to be connected with the potential harm of any confidentiality breach.* ***However, as we will see, sensitivity is a larger concept than this and encompasses how the data were collected and what reasonable expectations a data subject might hold about what will happen to data about them****"* (emphasis added). Nonetheless, *t*he fact that a confidentiality agreement is drawn to the subjects' attention would not be sufficient per se to transform data relating to persons into personal data (it is clearly not relevant in scenarios where a confidentiality agreement is signed but the relevant data is not of an obviously confidential/private nature). This issue requires further research consideration.

resulting from particular personal data processing activity). That is, when considering secondary harm (relevant to all cases where re-identification from data is a pre-requisite for harm potentially befalling a data subject), the normal Means Test would still require assessment following legal guidance related to the same.

Very importantly, however, it is worth noting that the *QMUL* facts are unusual because FOIA rules allow disclosure of non-personal data *without* the ability to impose restrictions on the requester (such as that requested data not be made public). Whereas, normally, a data controller would be able to mitigate both risks of re-identification *and* harm in respect of the intended activity of data sharing through the imposition of disclosure limitations (or, ultimately, deciding not to disclose the data at all).[803]

B. Critical analysis of the Effects-based Approach's compatibility with data protection's twin aims applied to the non-personal data concept

The Effects-based Approach (as developed so far) acknowledges the significance of context, and is a dynamic assessment model insofar as it focuses on the likely effects that intended usages of data are likely to have upon data subjects.[804] Applied to the non-personal data concept, this shifts focus onto possibilities for mitigating harm resulting from particular personal data processing activity so a Relevant Effect is no longer (deemed) likely to flow from it. Practically, therefore, the Effects-based Approach encourages data controllers to carry out (risk of) harm-mitigation techniques bespoke to the circumstances regarding each future processing activity type intended for the modified data. In other words, it discounts clearly any assumption that data controllers can anonymise-and-then-forget data.

Proponents of a dynamic re-identification risk-based model of personal data may also claim this strength. However, by comparison, the Effects-based Approach benefits from a more direct link to protecting rights (including privacy-related rights) by focusing on - not just reducing the risk of re-identification occurring (in the case of limiting the potential for secondary harm) - but also reducing

---

[803] The Code, p.37: *"[i]t is important to draw a distinction between the publication of anonymised data to the world at large and limited access".* When assessing whether data are considered personal and which data protection obligations should apply, respect should be had not only for technical de--identification but also for administrative and legal safeguards that make re-identification unlikely. Administrative safeguards include data security policies, access limits, employee training, data segregation guidelines, and data deletion practices that aim to stop confidential information from leaking. Legal controls include contractual terms that restrict how service providers or business associates may handle, use or share information, penalties for contractual breaches, and auditing rights to ensure compliance.

[804] In this sense, it is similar to certain re-identificatory risk-based models. See, e.g. comments by the Future of Privacy Forum (2014, p.8): "*whether a specific anonymization practice is appropriate will depend on the circumstances. When anonymizing data, organizations* ***should assess the risks*** *that the data could be re-identified* ***given the nature of the data, the context in which the data will be used, and the resources available to those with access to the data***" (emphasis added).

the risks of all harm that might otherwise ensue from processing a particular piece of data in a particular context. In other words, when organisations adjust their plans for processing usage of modified data, they do this directly to mitigate the risk of privacy harm associated with the use of that data, as this is the primary focus of the Effects-based Approach (rather than reducing re-identification risk as such except as a means to this overall end).

To this end, the framing language of the Effects-based Approach is not only more suited for those considering anonymisation but also more intuitive. For example, in the TGN, the ICO states:

> A significant consideration here is that as long as the first company have appropriate security in place there is little or no chance that any other person who might have access to the coded records would be able to link an individual by name and or address to a particular record. **In such circumstances the chances of an individual suffering detriment are negligible**.[805] (emphasis added)

Moreover, this overall focus on limiting the risk of (appreciable) harm befalling the individual to whom the data relates – if possible, or else accept data protection obligations – exists in two (plus) time periods. It applies before the modified data is shared (in respect of the intended processing activity of disclosure), but also after such disclosure has taken place (i.e. potentially ensuing upon further processing of that data by recipients in relation to whom data protection obligations may also become relevant).

There are other, additional (over and above those outlined in Chapter 4) strengths to the Effects-based Approach, such as that it avoids the confusion introduced by the GDPR about the significance of the new 'pseudonymisation' definition (see sub-section 5.2.3 above) and indeed its ramifications for the status of encrypted data.[806] Another practical advantage relates to the strong public interest in the processing of longitudinal research data that have been subject to key-coding techniques, especially to enable linkability of individual-level information across datasets to create longitudinal records. As described above, under a non-relativistic perspective on re-identifiability, if the data

---

[805] TGN, pp.28-9.

[806] Compare, Hon et al (2011, p.224) and their criticism of a strictly-objective stance to an identificatory-approach to personal data, illustrated through cloud computing hosting scenarios, because: *"excluding identification may be impossible. So too with SaaS 'passive' storage services, IaaS, or PaaS where, although stored unencrypted data are meant to be accessible only to the user, to investigate problems the provider's engineers need the ability to login to users' accounts or view stored data, and accordingly may see any identifying information therein. Similarly, where comprehensible 'personal data' shards remain temporarily on providers' equipment, pending automatic overwriting by other data after users decide to delete data or terminate accounts, at least where the provider can read shards marked for deletion. Therefore, currently, it seems some data stored by these cloud services must be treated as 'personal data'. This appears inevitable from WP136's focus on preventing identification, rather than assessing risks to privacy in context".* The only exception, they say, where this does not seem an unavoidable conclusion would be cloud services with total end-to-end strong encryption where the provider is unable to ever access stored encrypted data. For further criticisms of this stance as has been put forward by the WP, see also Hon et al (2014).

controller retains the raw data and key code, the modified data will remain personal data. The consequences of this interpretation may be to discourage de-identification by controllers increasing the privacy risk to trial participants, or it could push them to share less key-coded data with a chilling effect on the flow of such data. Under the Effects-based Approach, by comparison in that context, the relevant question to be asked is whether a processing activity in relation to the modified data – in particular, its potential disclosure to a third party - would have a Relevant Effect. To avoid that conclusion would require implementing safeguards to mitigate the risk of harm to the data subject, specifically to provide assurance that the means to reverse the key-coded data are controlled-tightly. For example, a research institute may have no means to access the key – say, it is was held by a sponsoring entity - and this may be enforced through an agreement with the sponsor prohibiting the sharing of keys under any circumstances or through organisational policies prohibiting such an exchange. Also for consideration under both approaches are safeguards to mitigate the risks of illegitimate access to information (e.g., physical and logical protection of databases, internal procedures adopted by data controllers, etc.).

In fact, many of such safeguards would be identical to those adopted under a relativist model of the (re-)identificatory-approach to personal data.[807] However, as explained, such model and the Effects-based Approach focus on different ultimate outcomes. For example, the Effects-based Approach would justify the putting into place of *on-going* safeguards to prevent relevant individuals being targeted by some attacker in a way that might cause them appreciable harm, which is especially concerning where the attacker's intentions are malicious. Alternatively, it could be accepted by the data controller that data protection rules will apply but then similar measures would also be required to put into place to fulfil data protection compliance obligations. Said otherwise, the greatest mitigatory incentive for data controllers is the possibility of the data protection regime applying. The conceptual marrying and inter-linking of these alternate possibilities as on-going considerations seems doctrinally more coherent under the Effects-based Approach, with the reality that either way data controllers cannot relinquish their responsibilities where data subjects' rights are at risk of being overridden to their appreciable detriment.[808]

---

[807] In other words, the data controller would also likely adopt appropriate measures of a similar type to ensure that the Means Test fails from the perspective of any third parties. As Knight & Stalla-Bourdillon point out (2016, pp.298-299): "i*f it is possible to isolate the raw datasets from the transformed datasets and put in place security measures, including technical and organisational measures, as well as legal obligations (essentially contractual obligations), so that the subsequent recipient of the transformed dataset will never have access to the raw dataset, the transformed dataset should be deemed as comprising data rendered anonymous at the very least in the hands of the subsequent recipient of the dataset".* According to the FTT in the *QMUL* judgement, this is because the possibility that data recipients would be able to re-identify the data subjects is remote, e.g. because they do not have access to the original data, and so the modified data is not personal data.

[808] Compare the proposal in Knight & Stalla-Bourdillon (2016, p.320) suggesting the exclusion of a certain number of recipients from the category of data controllers, in particular researchers.

For example, where longitudinal research studies take place using key-coded data, it may not be possible to go back to the original data subject to obtain their consent for further processing activities to be carried out on such data. However, if data controllers take necessary steps to safeguard such individuals from the risk of appreciable harm despite going ahead (where an alternate legal basis is not available), then this seems acceptable intuitively to ensure that data protection rules should not apply (as it goes further than just requiring that such data subjects are not re-identifiable).[809]

There are also weaknesses of the Effects-based Approach that cannot be ignored. Without repeating the arguments made in Chapter 4, suffice to say that the ICO has acknowledged explicitly problems around legal certainty in predicting likely effects within the Code.[810] More pertinently in the context of discussing the concept of non-personal data, the Effects-based Approach could be criticised for adding an additional layer of analysis over and above a (re-)identificatory-approach to personal data.[811] Moreover, in case of the disclosure of modified data to third parties, a data controller must think beyond one particular processing activity to further activities that recipients might carry out and their likely harm to the original data subject (i.e. indirect harm). On the other hand, opting for a bi-factorial test arguably reduces legal uncertainty which bedevils sole reliance on the (re)identificatory-approach. The essential difference is that the additional factor – the harm analysis - more explicitly synergises with and therefore safeguards the core purposes of data protection.

C. The GDPR

As mentioned, there is some (re-)identificatory approach-related confusion around the GDPR's references to pseudonymisation (its definition in Article 4(5) compared with other references, notably in Recital 26). However, such references fit in with the general logic of the Effects-based Approach.

Specifically, the GDPR describes pseudonymisation as a privacy-enhancing procedure (exemplifying the 'data protection by design and by default' principle, Article 25) that can reduce privacy risks for

[809] Moreover, the Effects-based Approach is flexible enough as a model of harm prevention that it could be expanded (if justification were found) to encompass consideration of potential harm not just to the original data subjects but also to others that might be affected appreciably (e.g. to encompass consideration of harm that might be suffered by data subject groups). This proposition is considered in Chapter 6 above, and is a suitable candidate for further research.
[810] The Code, p.25: *"it is good practice when releasing anonymised data to try to assess…what the consequences of re-identification are likely to be, if any, for the data subject concerned. Of course this can be difficult to assess in practice and a member of the public's sensitivity may be different from yours. For example, the disclosure of the address of a person on a witness protection scheme could be far more consequential than would usually be the case".*
[811] That is, whether the processing activity would reasonably likely cause appreciable harm to the data subject.

data subjects when data relating to them that has been subject to this procedure are processed. For example, Recital 28 recognises that "*[t]he application of pseudonymisation to personal data can reduce the risks for the data subjects concerned and help controllers and processors meet their data protection obligations…[while] not intended to preclude any other measures of data protection*".[812] To this end, the GDPR seems to provide legal incentives for data controllers to use pseudonymisation.

Moreover, Recital 75 highlights a link between the generic notion of privacy risk and harm. It refers to the risks to the rights and freedoms of natural persons "*of varying likelihood and severity"* that may result from personal data processing - that in turn "*could lead to physical, material or non-material damage*" - including consequent to the "*unauthorised reversal of pseudonymisation*".[813]

D. Appendix conclusion

This Appendix has developed the analysis in Chapters 3 and 4, and acts as a preface to the analysis in Part II of Chapter 5 and Chapter 6, by addressing sub-research questions 1-2 from a different perspective on the concept of personal data (i.e. what it means for data relating to a particular person no longer to be considered personal). This has involved an appraisal of whether – in deeming data for processing no longer subject to data protection law – its twin goals are likely better achieved either because:

- a relationship of identification can no longer be deemed to exist between the data and that person post-modification (it ceases to identify or be identifiable of that individual); or,

[812] Even though pseudonymisation is not held out as a fail-safe quick-fix to data protection law compliance. For example, in remarking that the GDPR encourages pseudonymisation that could be used (in turn) to encourage the application of big data analytics (European Commission (2017, Questions and Answers – Data Protection Reform Package. [online]), the European Commission also admits in the same breath that businesses should still "*be able to anticipate and inform individuals of the potential uses and benefits of big data - even if the exact specifics of the analysis are not yet known*"; something that is not incompatible with them allowing *"raw data to be retained for big data, while protecting the rules of individuals*".

[813] Recital 75, GDPR (in full): "***[t]he risk to the rights and freedoms of natural persons,*** *of varying likelihood and severity, may result from personal data processing which could lead to physical, material or non-material damage, in particular: where the processing may give rise to 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; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data; where personal data are processed which reveal racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, and the processing of genetic data, data concerning health or data concerning sex life or criminal convictions and offences or related security measures; where personal aspects are evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles; where personal data of vulnerable natural persons, in particular of children, are processed; or where processing involves a large amount of personal data and affects a large number of data subjects"* (emphasis added).

- a relationship of effects can no longer be deemed to exist between the data and that person (an intended processing activity would cease to have a Relevant Effect on that person, whereas pre-modification in respect of the same activity, *ceteris paribus*, a Relevant Effect would have been so construed).[814]

Hence, one approach confines the notion of non-personal data to the end-goal of ensuring that data subjects can no longer be identified from it, in relation to which this Part has demonstrated that pockets of legal uncertainty remain regarding what constitutes adequate legal de-identification of personal data (e.g. around acceptable levels of residual risk, and relevant factors for consideration). Partly this problem stems from terminological confusions, such as around the terms 'anonymous' and 'pseudonymous' (indeed, these terms are often used interchangeably and in an overlapping fashion.)[815] Such confusion reflects a steadfast attachment to static notions of anonymous data and pseudonymous data (or, indeed, encrypted data) as fixed qualities that become inherent or reified in data once-and-for-all-time.[816]

The second approach seeks to ensure that data protection rules would *only* not apply where the processing of information relating to individuals would no longer be reasonably likely to negatively affect (harm) them. Types of harm that may be considered would include, not only those that might flow secondarily to an individual if they were re-identified from the modified data, but also any arising independent of that possibility. Thus, while still intertwined with interpretations of re-identifiability as a risk-based prospect for assessment, this issue is given less prominence under the Effects-based Approach as the focus is on harm mitigation.[817]

Even where a dynamic (and relativist) concept of non-personal data is adopted under a (re-)identificatory model, the GDPR's focus appears more rooted on mitigating privacy risks

---

[814] In other words, whereas pre-modification the same piece of data - considered in the context of applying the same intended processing activity - would have been so deemed likely.

[815] Compare, e.g., Stalla-Bourdillon (2016, A call for a common techno-legal language to speak about anonymisation, pseudonymisation, de-identification… Could this be one of the biggest challenges brought about by the GDPR? [online]).

[816] Compare Wong (2013, p.83): "*[t]here is a need to clarify the concept of personal data. The question to be considered is whether a relative approach should be adopted in the context of encoded or pseudonymised data? The concept of personal data is broad to cover identified or identifiable persons. There is uncertainty about what "identifiable" covers? The question is not whether anyone can identify that data X belongs to person A. Indeed, it is arguable that the data controller cannot exercise his duties under the relevant data protection laws, if he/she cannot identify that data X belongs to person A.* ***I would even contend that if person B cannot identify data X then it is not personal data, but if person A can, then it is personal data. Data X does not cease to be personal data on the basis that person B cannot identify that X is personal data"*** (emphasis added). See also the European Parliament's legislative resolution of 12 March 2014 (P7_TA(2014)0212). It proposed that a person may be subjected to profiling that leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject provided that the conditions set out in Article 20(2) are met; however, profiling based solely on the processing of pseudonymous data would be presumed not to significantly affect the interests, rights or freedoms of the data subject (Recital 58a).

[817] For example, appreciable harm might be deemed reasonably liable to result from analysing data in order to make decisions about singled out individuals whose 'real-world' identities are unknown, regardless of what it is concluded about the capacity for identifying such individuals.

potentially harming the data subject, in discussing the application of pseudonymisation as a technique. Said otherwise, while issues of (re-)identifiability assessment remain confused under the GDPR, there is more certainty about the need to reduce the risk of harm to individuals (whether in a jurisdictional *or* substantive sense when the data protection regime applies, and in fact discussion of the two possibilities overlap) to ensure a high level of data protection rights-related protection. That marriage of alternatives provides a better balance between the twin objectives also in highlighting their joint promotion as compatible and part of the same narrative of regulatory compliance.

However, this does leave us in the position that the hypothesis put forward by Hon et al might be deemed *as effective* as the Effects-based Approach and a good deal easier to implement when it comes to the concept of non-personal data. As a reminder, they argue that – while the definition of personal data "*should be based on a realistic risk of identification*" – the "*applicability of data protection rules should be based on risk of harm and its likely severity*" (see fn.557-558 above).

Redirecting the reader back to Chapter 5, Part II, this issue is addressed, along with how it might be possible, therefore, to make an effects-centric analysis more attractive through remodelling (in particular, by considering an assessment framework that could be employed by data controllers when deciding whether personal data exists in particular circumstances). Moreover, there is consideration of incentives and development of the argument that an effects-centric analysis might provide additional benefits in encouraging data controllers to engage with jurisdictional issues in more than a perfunctory one-off manner (as is often the case now when it comes to considering data from which indirect identifiers have been removed).

Appendix 5 – The Vertical Block Exemption

COMMISSION REGULATION (EU) No 330/2010 of 20 April 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of vertical agreements and concerted practices

(Text with EEA relevance)

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Regulation No 19/65/EEC of the Council of 2 March 1965 on the application of Article 85(3) of the Treaty to certain categories of agreements and concerted practices [(1)], and in particular Article 1 thereof,

Having published a draft of this Regulation,

After consulting the Advisory Committee on Restrictive Practices and Dominant Positions,

Whereas:

(1)Regulation No 19/65/EEC empowers the Commission to apply Article 101(3) of the Treaty on the Functioning of the European Union [(2)] by regulation to certain categories of vertical agreements and corresponding concerted practices falling within Article 101(1) of the Treaty.

(2)Commission Regulation (EC) No 2790/1999 of 22 December 1999 on the application of Article 81(3) of the Treaty to categories of vertical agreements and concerted practices [(3)] defines a category of vertical agreements which the Commission regarded as normally satisfying the conditions laid down in Article 101(3) of the Treaty. In view of the overall positive experience with the application of that Regulation, which expires on 31 May 2010, and taking into account further experience acquired since its adoption, it is appropriate to adopt a new block exemption regulation.

(3)The category of agreements which can be regarded as normally satisfying the conditions laid down in Article 101(3) of the Treaty includes vertical agreements for the purchase or sale of goods or services where those agreements are concluded between non-competing undertakings, between certain competitors or by certain associations of retailers of goods. It also includes vertical agreements containing ancillary provisions on the assignment or use of intellectual property rights. The term 'vertical agreements' should include the corresponding concerted practices.

(4)For the application of Article 101(3) of the Treaty by regulation, it is not necessary to define those vertical agreements which are capable of falling within Article 101(1) of the Treaty. In the individual assessment of agreements under Article 101(1) of the Treaty, account has to be taken of several factors, and in particular the market structure on the supply and purchase side.

(5)The benefit of the block exemption established by this Regulation should be limited to vertical agreements for which it can be assumed with sufficient certainty that they satisfy the conditions of Article 101(3) of the Treaty.

(6)Certain types of vertical agreements can improve economic efficiency within a chain of production or distribution by facilitating better coordination between the participating undertakings. In particular, they can lead to a reduction in the transaction and distribution costs of the parties and to an optimisation of their sales and investment levels.

(7)The likelihood that such efficiency-enhancing effects will outweigh any anti-competitive effects due to restrictions contained in vertical agreements depends on the degree of market power of the parties to the agreement and, therefore, on the extent to which those undertakings face competition from other suppliers of goods or services regarded by their customers as interchangeable or substitutable for one another, by reason of the products' characteristics, their prices and their intended use.

(8)It can be presumed that, where the market share held by each of the undertakings party to the agreement on the relevant market does not exceed 30 %, vertical agreements which do not contain certain types of severe restrictions of competition generally lead to an improvement in production or distribution and allow consumers a fair share of the resulting benefits.

(9)Above the market share threshold of 30 %, there can be no presumption that vertical agreements falling within the scope of Article 101(1) of the Treaty will usually give rise to objective advantages of such a character and size as to compensate for the disadvantages which they create for competition. At the same time, there is no presumption that those vertical agreements are either caught by Article 101(1) of the Treaty or that they fail to satisfy the conditions of Article 101(3) of the Treaty.

(10)This Regulation should not exempt vertical agreements containing restrictions which are likely to restrict competition and harm consumers or which are not indispensable to the attainment of the efficiency-enhancing effects. In particular, vertical agreements containing certain types of severe restrictions of competition such as minimum and fixed resale-prices, as well as certain types of territorial protection, should be excluded from the benefit of the block exemption established by this Regulation irrespective of the market share of the undertakings concerned.

(11)In order to ensure access to or to prevent collusion on the relevant market, certain conditions should be attached to the block exemption. To this end, the exemption of non-compete obligations should be limited to obligations which do not exceed a defined duration. For the same reasons, any direct or indirect obligation causing the members of a selective distribution system not to sell the brands of particular competing suppliers should be excluded from the benefit of this Regulation.

(12)The market-share limitation, the non-exemption of certain vertical agreements and the conditions provided for in this Regulation normally ensure that the agreements to which the block exemption applies do not enable the participating undertakings to eliminate competition in respect of a substantial part of the products in question.

(13)The Commission may withdraw the benefit of this Regulation, pursuant to Article 29(1) of Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty [(4)], where it finds in a particular case that an agreement to which the exemption provided for in this Regulation applies nevertheless has effects which are incompatible with Article 101(3) of the Treaty.

(14)The competition authority of a Member State may withdraw the benefit of this Regulation pursuant to Article 29(2) of Regulation (EC) No 1/2003 in respect of the territory of that Member State, or a part thereof where, in a particular case, an agreement to which the exemption provided for in this Regulation applies nevertheless has effects which are incompatible with Article 101(3) of the Treaty in the territory of that Member State, or in a part thereof, and where such territory has all the characteristics of a distinct geographic market.

(15)In determining whether the benefit of this Regulation should be withdrawn pursuant to Article 29 of Regulation (EC) No 1/2003, the anti-competitive effects that may derive from the existence of parallel networks of vertical agreements that have similar effects which significantly restrict access to a relevant market or competition therein are of particular importance. Such cumulative effects may for example arise in the case of selective distribution or non compete obligations.

(16)In order to strengthen supervision of parallel networks of vertical agreements which have similar anti-competitive effects and which cover more than 50 % of a given market, the Commission may by regulation declare this Regulation inapplicable to vertical agreements containing specific restraints relating to the market concerned, thereby restoring the full application of Article 101 of the Treaty to such agreements,

HAS ADOPTED THIS REGULATION:

Appendices

Article 1

Definitions

1. For the purposes of this Regulation, the following definitions shall apply:

(a)'vertical agreement' means an agreement or concerted practice entered into between two or more undertakings each of which operates, for the purposes of the agreement or the concerted practice, at a different level of the production or distribution chain, and relating to the conditions under which the parties may purchase, sell or resell certain goods or services;

(b)'vertical restraint' means a restriction of competition in a vertical agreement falling within the scope of Article 101(1) of the Treaty;

(c)'competing undertaking' means an actual or potential competitor; 'actual competitor' means an undertaking that is active on the same relevant market; 'potential competitor' means an undertaking that, in the absence of the vertical agreement, would, on realistic grounds and not just as a mere theoretical possibility, in case of a small but permanent increase in relative prices be likely to undertake, within a short period of time, the necessary additional investments or other necessary switching costs to enter the relevant market;

(d)'non-compete obligation' means any direct or indirect obligation causing the buyer not to manufacture, purchase, sell or resell goods or services which compete with the contract goods or services, or any direct or indirect obligation on the buyer to purchase from the supplier or from another undertaking designated by the supplier more than 80 % of the buyer's total purchases of the contract goods or services and their substitutes on the relevant market, calculated on the basis of the value or, where such is standard industry practice, the volume of its purchases in the preceding calendar year;

(e)'selective distribution system' means a distribution system where the supplier undertakes to sell the contract goods or services, either directly or indirectly, only to distributors selected on the basis of specified criteria and where these distributors undertake not to sell such goods or services to unauthorised distributors within the territory reserved by the supplier to operate that system;

(f)'intellectual property rights' includes industrial property rights, know how, copyright and neighbouring rights;

(g)'know-how' means a package of non-patented practical information, resulting from experience and testing by the supplier, which is secret, substantial and identified: in this context, 'secret' means that the know-how is not generally known or easily accessible; 'substantial' means that the know-how is significant and useful to the buyer for the use, sale or resale of the contract

goods or services; 'identified' means that the know-how is described in a sufficiently comprehensive manner so as to make it possible to verify that it fulfils the criteria of secrecy and substantiality;

(h)'buyer' includes an undertaking which, under an agreement falling within Article 101(1) of the Treaty, sells goods or services on behalf of another undertaking;

(i)'customer of the buyer' means an undertaking not party to the agreement which purchases the contract goods or services from a buyer which is party to the agreement.

2. For the purposes of this Regulation, the terms 'undertaking', 'supplier' and 'buyer' shall include their respective connected undertakings.

'Connected undertakings' means:

(a)undertakings in which a party to the agreement, directly or indirectly:

(i) has the power to exercise more than half the voting rights, or

(ii)has the power to appoint more than half the members of the supervisory board, board of management or bodies legally representing the undertaking, or

(iii) has the right to manage the undertaking's affairs;

(b)undertakings which directly or indirectly have, over a party to the agreement, the rights or powers listed in point (a);

(c)undertakings in which an undertaking referred to in point (b) has, directly or indirectly, the rights or powers listed in point (a);

(d)undertakings in which a party to the agreement together with one or more of the undertakings referred to in points (a), (b) or (c), or in which two or more of the latter undertakings, jointly have the rights or powers listed in point (a);

(e)undertakings in which the rights or the powers listed in point (a) are jointly held by:

(i)parties to the agreement or their respective connected undertakings referred to in points (a) to (d), or

(ii)one or more of the parties to the agreement or one or more of their connected undertakings referred to in points (a) to (d) and one or more third parties.

Article 2

Exemption

1. Pursuant to Article 101(3) of the Treaty and subject to the provisions of this Regulation, it is hereby declared that Article 101(1) of the Treaty shall not apply to vertical agreements.

This exemption shall apply to the extent that such agreements contain vertical restraints.

2. The exemption provided for in paragraph 1 shall apply to vertical agreements entered into between an association of undertakings and its members, or between such an association and its suppliers, only if all its members are retailers of goods and if no individual member of the association, together with its connected undertakings, has a total annual turnover exceeding EUR 50 million. Vertical agreements entered into by such associations shall be covered by this Regulation without prejudice to the application of Article 101 of the Treaty to horizontal agreements concluded between the members of the association or decisions adopted by the association.

3. The exemption provided for in paragraph 1 shall apply to vertical agreements containing provisions which relate to the assignment to the buyer or use by the buyer of intellectual property rights, provided that those provisions do not constitute the primary object of such agreements and are directly related to the use, sale or resale of goods or services by the buyer or its customers. The exemption applies on condition that, in relation to the contract goods or services, those provisions do not contain restrictions of competition having the same object as vertical restraints which are not exempted under this Regulation.

4. The exemption provided for in paragraph 1 shall not apply to vertical agreements entered into between competing undertakings. However, it shall apply where competing undertakings enter into a non-reciprocal vertical agreement and:

(a)the supplier is a manufacturer and a distributor of goods, while the buyer is a distributor and not a competing undertaking at the manufacturing level; or

(b)the supplier is a provider of services at several levels of trade, while the buyer provides its goods or services at the retail level and is not a competing undertaking at the level of trade where it purchases the contract services.

5. This Regulation shall not apply to vertical agreements the subject matter of which falls within the scope of any other block exemption regulation, unless otherwise provided for in such a regulation.

Article 3

Market share threshold

1. The exemption provided for in Article 2 shall apply on condition that the market share held by the supplier does not exceed 30 % of the relevant market on which it sells the contract goods or services and the market share held by the buyer does not exceed 30 % of the relevant market on which it purchases the contract goods or services.

2. For the purposes of paragraph 1, where in a multi party agreement an undertaking buys the contract goods or services from one undertaking party to the agreement and sells the contract goods or services to another undertaking party to the agreement, the market share of the first undertaking must respect the market share threshold provided for in that paragraph both as a buyer and a supplier in order for the exemption provided for in Article 2 to apply.

Article 4

Restrictions that remove the benefit of the block exemption — hardcore restrictions

The exemption provided for in Article 2 shall not apply to vertical agreements which, directly or indirectly, in isolation or in combination with other factors under the control of the parties, have as their object:

(a)the restriction of the buyer's ability to determine its sale price, without prejudice to the possibility of the supplier to impose a maximum sale price or recommend a sale price, provided that they do not amount to a fixed or minimum sale price as a result of pressure from, or incentives offered by, any of the parties;

(b)the restriction of the territory into which, or of the customers to whom, a buyer party to the agreement, without prejudice to a restriction on its place of establishment, may sell the contract goods or services, except:

(i)the restriction of active sales into the exclusive territory or to an exclusive customer group reserved to the supplier or allocated by the supplier to another buyer, where such a restriction does not limit sales by the customers of the buyer,

(ii)the restriction of sales to end users by a buyer operating at the wholesale level of trade,

(iii)the restriction of sales by the members of a selective distribution system to unauthorised distributors within the territory reserved by the supplier to operate that system, and

(iv)the restriction of the buyer's ability to sell components, supplied for the purposes of incorporation, to customers who would use them to manufacture the same type of goods as those produced by the supplier;

(c)the restriction of active or passive sales to end users by members of a selective distribution system operating at the retail level of trade, without prejudice to the possibility of prohibiting a member of the system from operating out of an unauthorised place of establishment;

(d)the restriction of cross-supplies between distributors within a selective distribution system, including between distributors operating at different level of trade;

(e)the restriction, agreed between a supplier of components and a buyer who incorporates those components, of the supplier's ability to sell the components as spare parts to end-users or to repairers or other service providers not entrusted by the buyer with the repair or servicing of its goods.

Article 5

Excluded restrictions

1. The exemption provided for in Article 2 shall not apply to the following obligations contained in vertical agreements:

(a)any direct or indirect non-compete obligation, the duration of which is indefinite or exceeds five years;

(b)any direct or indirect obligation causing the buyer, after termination of the agreement, not to manufacture, purchase, sell or resell goods or services;

(c)any direct or indirect obligation causing the members of a selective distribution system not to sell the brands of particular competing suppliers.

For the purposes of point (a) of the first subparagraph, a non-compete obligation which is tacitly renewable beyond a period of five years shall be deemed to have been concluded for an indefinite duration.

2. By way of derogation from paragraph 1(a), the time limitation of five years shall not apply where the contract goods or services are sold by the buyer from premises and land owned by the supplier or leased by the supplier from third parties not connected with the buyer, provided that the duration of the non-compete obligation does not exceed the period of occupancy of the premises and land by the buyer.

3. By way of derogation from paragraph 1(b), the exemption provided for in Article 2 shall apply to any direct or indirect obligation causing the buyer, after termination of the agreement, not to manufacture, purchase, sell or resell goods or services where the following conditions are fulfilled:

(a)the obligation relates to goods or services which compete with the contract goods or services;

(b)the obligation is limited to the premises and land from which the buyer has operated during the contract period;

(c)the obligation is indispensable to protect know-how transferred by the supplier to the buyer;

(d)the duration of the obligation is limited to a period of one year after termination of the agreement.

Paragraph 1(b) is without prejudice to the possibility of imposing a restriction which is unlimited in time on the use and disclosure of know-how which has not entered the public domain.

Article 6

Non-application of this Regulation

Pursuant to Article 1a of Regulation No 19/65/EEC, the Commission may by regulation declare that, where parallel networks of similar vertical restraints cover more than 50 % of a relevant market, this Regulation shall not apply to vertical agreements containing specific restraints relating to that market.

Article 7

Application of the market share threshold

For the purposes of applying the market share thresholds provided for in Article 3 the following rules shall apply:

(a)the market share of the supplier shall be calculated on the basis of market sales value data and the market share of the buyer shall be calculated on the basis of market purchase value data. If market sales value or market purchase value data are not available, estimates based on other reliable market information, including market sales and purchase volumes, may be used to establish the market share of the undertaking concerned;

(b)the market shares shall be calculated on the basis of data relating to the preceding calendar year;

(c)the market share of the supplier shall include any goods or services supplied to vertically integrated distributors for the purposes of sale;

(d)if a market share is initially not more than 30 % but subsequently rises above that level without exceeding 35 %, the exemption provided for in Article 2 shall continue to apply for a period of two consecutive calendar years following the year in which the 30 % market share threshold was first exceeded;

(e)if a market share is initially not more than 30 % but subsequently rises above 35 %, the exemption provided for in Article 2 shall continue to apply for one calendar year following the year in which the level of 35 % was first exceeded;

(f)the benefit of points (d) and (e) may not be combined so as to exceed a period of two calendar years;

(g)the market share held by the undertakings referred to in point (e) of the second subparagraph of Article 1(2) shall be apportioned equally to each undertaking having the rights or the powers listed in point (a) of the second subparagraph of Article 1(2).

Article 8

Application of the turnover threshold

1. For the purpose of calculating total annual turnover within the meaning of Article 2(2), the turnover achieved during the previous financial year by the relevant party to the vertical agreement and the turnover achieved by its connected undertakings in respect of all goods and services, excluding all taxes and other duties, shall be added together. For this purpose, no account shall be taken of dealings between the party to the vertical agreement and its connected undertakings or between its connected undertakings.

2. The exemption provided for in Article 2 shall remain applicable where, for any period of two consecutive financial years, the total annual turnover threshold is exceeded by no more than 10 %.

Article 9

Transitional period

The prohibition laid down in Article 101(1) of the Treaty shall not apply during the period from 1 June 2010 to 31 May 2011 in respect of agreements already in force on 31 May 2010 which do not satisfy the conditions for exemption provided for in this Regulation but which, on 31 May 2010, satisfied the conditions for exemption provided for in Regulation (EC) No 2790/1999.

Article 10

Period of validity

This Regulation shall enter into force on 1 June 2010.

It shall expire on 31 May 2022.

This Regulation shall be binding in its entirety and directly applicable in all Member States.

Done at Brussels, 20 April 2010.

For the Commission

The President

José Manuel BARROSO

([1]) OJ 36, 6.3.1965, p. 533.

([2]) With effect from 1 December 2009, Article 81 of the EC Treaty has become Article 101 of the Treaty on the Functioning of the European Union. The two Articles are, in substance, identical. For the purposes of this Regulation, references to Article 101 of the Treaty on the Functioning of the European Union should be understood as references to Article 81 of the EC Treaty where appropriate.

([3]) OJ L 336, 29.12.1999, p. 21.

([4]) OJ L 1, 4.1.2003, p. 1.

Appendix 6 – Extracts from the European Commission Guidelines on Vertical Restraints

Guidelines on Vertical Restraints (emphasis added by the thesis author in underline below to highlight analogous relevance to an effects-centric analysis under data protection law)

(Text with EEA relevance)

Official Journal of the EU 2010/C 130/01

TABLE OF CONTENTS

I. INTRODUCTION

1. Purpose of the Guidelines

(1)These Guidelines set out the principles for the assessment of vertical agreements under Article 101 of the Treaty on the Functioning of the European Union [(1)] (hereinafter 'Article 101') [(2)]. Article 1(1)(a) of Commission Regulation (EU) No 330/2010 of 20 April 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of vertical agreements and concerted practices [(3)] (hereinafter referred to as the 'Block Exemption Regulation') (see paragraphs (24) to (46)) defines the term 'vertical agreement'. These Guidelines are without prejudice to the possible parallel application of Article 102 of the Treaty on the Functioning of the European Union (hereinafter 'Article 102') to vertical agreements. These Guidelines are structured in the following way:

—Section II (paragraphs (8) to (22)) describes vertical agreements which generally fall outside Article 101(1);

—Section III (paragraphs (23) to (73)) clarifies the conditions for the application of the Block Exemption Regulation;

—Section IV (paragraphs (74) to (85)) describes the principles concerning the withdrawal of the block exemption and the disapplication of the Block Exemption Regulation;

—Section V (paragraphs (86) to (95)) provides guidance on how to define the relevant market and calculate market shares;

—Section VI (paragraphs (96) to (229)) describes the general framework of analysis and the enforcement policy of the Commission in individual cases concerning vertical agreements.

(2)Throughout these Guidelines, the analysis applies to both goods and services, although certain vertical restraints are mainly used in the distribution of goods. Similarly, vertical agreements can be concluded for intermediate and final goods and services. Unless otherwise stated, the analysis and arguments in these Guidelines apply to all types of goods and services and to all levels of trade. Thus, the term 'products' includes both goods and services. The terms 'supplier' and 'buyer' are used for all levels of trade. The Block Exemption Regulation and these Guidelines do not apply to agreements with final consumers where the latter are not undertakings, since Article 101 only applies to agreements between undertakings.

(3)By issuing these Guidelines, the Commission aims to help companies conduct their own assessment of vertical agreements under EU competition rules. The standards set forth in these Guidelines cannot be applied mechanically, but must be applied with due consideration for the specific circumstances of each case. Each case must be evaluated in the light of its own facts.

(4)These Guidelines are without prejudice to the case-law of the General Court and the Court of Justice of the European Union concerning the application of Article 101 to vertical agreements. The Commission will continue to monitor the operation of the Block Exemption Regulation and Guidelines based on market information from stakeholders and national competition authorities and may revise this notice in the light of future developments and of evolving insight.

2. Applicability of Article 101 to vertical agreements

(5)Article 101 applies to vertical agreements that may affect trade between Member States and that prevent, restrict or distort competition ('vertical restraints') [(4)]. Article 101 provides a legal framework for the assessment of vertical restraints, which takes into consideration the distinction between anti-competitive and pro-competitive effects. Article 101(1) prohibits those agreements which appreciably restrict or distort competition, while Article 101(3) exempts those agreements which confer sufficient benefits to outweigh the anti-competitive effects [(5)].

(6)For most vertical restraints, competition concerns can only arise if there is insufficient competition at one or more levels of trade, that is, if there is some degree of market power at the level of the supplier or the buyer or at both levels. Vertical restraints are generally less harmful than horizontal restraints and may provide substantial scope for efficiencies.

(7)The objective of Article 101 is to ensure that undertakings do not use agreements – in this context, vertical agreements – to restrict competition on the market to the detriment of consumers. Assessing vertical restraints is also important in the context of the wider objective of achieving an integrated internal market. Market integration enhances competition in the European Union. Companies should not be allowed to re-establish private barriers between Member States where State barriers have been successfully abolished.

II. VERTICAL AGREEMENTS WHICH GENERALLY FALL OUTSIDE THE SCOPE OF ARTICLE 101(1)

1. Agreements of minor importance and SMEs

(8)Agreements that are not capable of appreciably affecting trade between Member States or of appreciably restricting competition by object or effect do not fall within the scope of Article 101(1). The Block Exemption Regulation applies only to agreements falling within the scope of application of Article 101(1). These Guidelines are without prejudice to the application of Commission Notice on agreements of minor importance which do not appreciably restrict competition under Article 81(1) of the Treaty establishing the European Community (de minimis) [(6)] or any future de minimis notice.

(9)Subject to the conditions set out in the de minimis notice concerning hardcore restrictions and cumulative effect issues, vertical agreements entered into by non-competing undertakings whose individual market share on the relevant market does not exceed 15 % are generally considered to fall outside the scope of Article 101(1) [(7)]. There is no presumption that vertical agreements concluded by undertakings having more than 15 % market share automatically infringe Article 101(1). Agreements between undertakings whose market share exceeds the 15 % threshold may still not have an appreciable effect on trade between Member States or may not constitute an appreciable restriction of competition [(8)]. Such agreements need to be assessed in their legal and economic context. The criteria for the assessment of individual agreements are set out in paragraphs (96) to (229).

(10)As regards hardcore restrictions referred to in the de minimis notice, Article 101(1) may apply below the 15 % threshold, provided that there is an appreciable effect on trade between Member States and on competition. The applicable case-law of the Court of Justice and the General Court is relevant in this respect [(9)]. Reference is also made to the possible need to assess positive and negative effects of hardcore restrictions as described in particular in paragraph (47) of these Guidelines.

(11)In addition, the Commission considers that, subject to cumulative effect and hardcore restrictions, vertical agreements between small and medium-sized undertakings as defined in the Annex to Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises [(10)] are rarely capable of appreciably affecting trade between Member States or of appreciably restricting competition within the meaning of Article 101(1), and therefore generally fall outside the scope of Article 101(1). In cases where such agreements nonetheless meet the conditions for the application of Article 101(1), the Commission will normally refrain from opening proceedings for lack of sufficient interest for the

European Union unless those undertakings collectively or individually hold a dominant position in a substantial part of the internal market.

[…]

III. APPLICATION OF THE BLOCK EXEMPTION REGULATION

1. Safe harbour created by the Block Exemption Regulation

(23)For most vertical restraints, competition concerns can only arise if there is insufficient competition at one or more levels of trade, that is, if there is some degree of market power at the level of the supplier or the buyer or at both levels. Provided that they do not contain hardcore restrictions of competition, which are restrictions of competition by object, the Block Exemption Regulation creates a presumption of legality for vertical agreements depending on the market share of the supplier and the buyer. Pursuant to Article 3 of the Block Exemption Regulation, it is the supplier's market share on the market where it sells the contract goods or services and the buyer's market share on the market where it purchases the contract goods or services which determine the applicability of the block exemption. In order for the block exemption to apply, the supplier's and the buyer's market share must each be 30 % or less. Section V of these Guidelines provides guidance on how to define the relevant market and calculate the market shares. Above the market share threshold of 30 %, there is no presumption that vertical agreements fall within the scope of Article 101(1) or fail to satisfy the conditions of Article 101(3) but there is also no presumption that vertical agreements falling within the scope of Article 101(1) will usually satisfy the conditions of Article 101(3).

[…]

IV. WITHDRAWAL OF THE BLOCK EXEMPTION AND DISAPPLICATION OF THE BLOCK EXEMPTION REGULATION

1. Withdrawal procedure

(74)The presumption of legality conferred by the Block Exemption Regulation may be withdrawn where a vertical agreement, considered either in isolation or in conjunction with similar agreements enforced by competing suppliers or buyers, comes within the scope of Article 101(1) and does not fulfil all the conditions of Article 101(3).

(75)The conditions of Article 101(3) may in particular not be fulfilled when access to the relevant market or competition therein is significantly restricted by the cumulative effect of parallel networks of similar vertical agreements practised by competing suppliers or buyers. Parallel networks of vertical agreements are to be regarded as similar if they contain restraints producing similar effects on the market. Such a situation may arise for example when, on a given

market, certain suppliers practise purely qualitative selective distribution while other suppliers practise quantitative selective distribution. Such a situation may also arise when, on a given market, the cumulative use of qualitative criteria forecloses more efficient distributors. In such circumstances, the assessment must take account of the anti-competitive effects attributable to each individual network of agreements. Where appropriate, withdrawal may concern only a particular qualitative criterion or only the quantitative limitations imposed on the number of authorised distributors.

(76)Responsibility for an anti-competitive cumulative effect can only be attributed to those undertakings which make an appreciable contribution to it. Agreements entered into by undertakings whose contribution to the cumulative effect is insignificant do not fall under the prohibition provided for in Article 101(1) [(33)] and are therefore not subject to the withdrawal mechanism. The assessment of such a contribution will be made in accordance with the criteria set out in paragraphs (128) to (229).

(77)Where the withdrawal procedure is applied, the Commission bears the burden of proof that the agreement falls within the scope of Article 101(1) and that the agreement does not fulfil one or several of the conditions of Article 101(3). A withdrawal decision can only have ex nunc effect, which means that the exempted status of the agreements concerned will not be affected until the date at which the withdrawal becomes effective.

(78)As referred to in recital 14 of the Block Exemption Regulation, the competition authority of a Member State may withdraw the benefit of the Block Exemption Regulation in respect of vertical agreements whose anti-competitive effects are felt in the territory of the Member State concerned or a part thereof, which has all the characteristics of a distinct geographic market. The Commission has the exclusive power to withdraw the benefit of the Block Exemption Regulation in respect of vertical agreements restricting competition on a relevant geographic market which is wider than the territory of a single Member State. When the territory of a single Member State, or a part thereof, constitutes the relevant geographic market, the Commission and the Member State concerned have concurrent competence for withdrawal.

[…]

VI. ENFORCEMENT POLICY IN INDIVIDUAL CASES

1. The framework of analysis

(96)Outside the scope of the block exemption, it is relevant to examine whether in the individual case the agreement falls within the scope of Article 101(1) and if so whether the conditions of Article 101(3) are satisfied. Provided that they do not contain restrictions of competition by object and in particular hardcore restrictions of competition, there is no presumption that

vertical agreements falling outside the block exemption because the market share threshold is exceeded fall within the scope of Article 101(1) or fail to satisfy the conditions of Article 101(3). Individual assessment of the likely effects of the agreement is required. Companies are encouraged to do their own assessment. Agreements that either do not restrict competition within the meaning of Article 101(1) or which fulfil the conditions of Article 101(3) are valid and enforceable. Pursuant to Article 1(2) of Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty [(38)] no notification needs to be made to benefit from an individual exemption under Article 101(3). In the case of an individual examination by the Commission, the latter will bear the burden of proof that the agreement in question infringes Article 101(1). The undertakings claiming the benefit of Article 101(3) bear the burden of proving that the conditions of that paragraph are fulfilled. When likely anti-competitive effects are demonstrated, undertakings may substantiate efficiency claims and explain why a certain distribution system is indispensable to bring likely benefits to consumers without eliminating competition, before the Commission decides whether the agreement satisfies the conditions of Article 101(3).

(97)The assessment of whether a vertical agreement has the effect of restricting competition will be made by comparing the actual or likely future situation on the relevant market with the vertical restraints in place with the situation that would prevail in the absence of the vertical restraints in the agreement. In the assessment of individual cases, the Commission will take, as appropriate, both actual and likely effects into account. For vertical agreements to be restrictive of competition by effect they must affect actual or potential competition to such an extent that on the relevant market negative effects on prices, output, innovation, or the variety or quality of goods and services can be expected with a reasonable degree of probability. The likely negative effects on competition must be appreciable [(39)]. Appreciable anticompetitive effects are likely to occur when at least one of the parties has or obtains some degree of market power and the agreement contributes to the creation, maintenance or strengthening of that market power or allows the parties to exploit such market power. Market power is the ability to maintain prices above competitive levels or to maintain output in terms of product quantities, product quality and variety or innovation below competitive levels for a not insignificant period of time. The degree of market power normally required for a finding of an infringement under Article 101(1) is less than the degree of market power required for a finding of dominance under Article 102.

(98)Vertical restraints are generally less harmful than horizontal restraints. The main reason for the greater focus on horizontal restraints is that such restraints may concern an agreement between competitors producing identical or substitutable goods or services. In such horizontal

relationships, the exercise of market power by one company (higher price of its product) may benefit its competitors. This may provide an incentive to competitors to induce each other to behave anti-competitively. In vertical relationships, the product of the one is the input for the other, in other words, the activities of the parties to the agreement are complementary to each other. The exercise of market power by either the upstream or downstream company would therefore normally hurt the demand for the product of the other. The companies involved in the agreement therefore usually have an incentive to prevent the exercise of market power by the other.

(99)Such self-restraining character should not, however, be over-estimated. When a company has no market power, it can only try to increase its profits by optimising its manufacturing and distribution processes, with or without the help of vertical restraints. More generally, because of the complementary role of the parties to a vertical agreement in getting a product on the market, vertical restraints may provide substantial scope for efficiencies. However, when an undertaking does have market power it can also try to increase its profits at the expense of its direct competitors by raising their costs and at the expense of its buyers and ultimately consumers by trying to appropriate some of their surplus. This can happen when the upstream and downstream company share the extra profits or when one of the two uses vertical restraints to appropriate all the extra profits.

1.1 Negative effects of vertical restraints

(100)The negative effects on the market that may result from vertical restraints which EU competition law aims at preventing are the following:

(a)anticompetitive foreclosure of other suppliers or other buyers by raising barriers to entry or expansion;

(b)softening of competition between the supplier and its competitors and/or facilitation of collusion amongst these suppliers, often referred to as reduction of inter-brand competition [(40)];

(c)softening of competition between the buyer and its competitors and/or facilitation of collusion amongst these competitors, often referred to as reduction of intra-brand competition if it concerns distributors' competition on the basis of the brand or product of the same supplier;

(d)the creation of obstacles to market integration, including, above all, limitations on the possibilities for consumers to purchase goods or services in any Member State they may choose.

(101)Foreclosure, softening of competition and collusion at the manufacturers' level may harm consumers in particular by increasing the wholesale prices of the products, limiting the choice of products, lowering their quality or reducing the level of product innovation. Foreclosure, softening of competition and collusion at the distributors' level may harm consumers in particular by increasing the retail prices of the products, limiting the choice of price-service combinations and distribution formats, lowering the availability and quality of retail services and reducing the level of innovation of distribution.

(102)On a market where individual distributors distribute the brand(s) of only one supplier, a reduction of competition between the distributors of the same brand will lead to a reduction of intra-brand competition between these distributors, but may not have a negative effect on competition between distributors in general. In such a case, if inter-brand competition is fierce, it is unlikely that a reduction of intra-brand competition will have negative effects for consumers.

(103)Exclusive arrangements are generally more anti-competitive than non-exclusive arrangements. Exclusive arrangements, whether by means of express contractual language or their practical effects, result in one party sourcing all or practically all of its demand from another party. For instance, under a non-compete obligation the buyer purchases only one brand. Quantity forcing, on the other hand, leaves the buyer some scope to purchase competing goods. The degree of foreclosure may therefore be less with quantity forcing.

(104)Vertical restraints agreed for non-branded goods and services are in general less harmful than restraints affecting the distribution of branded goods and services. Branding tends to increase product differentiation and reduce substitutability of the product, leading to a reduced elasticity of demand and an increased possibility to raise price. The distinction between branded and non-branded goods or services will often coincide with the distinction between intermediate goods and services and final goods and services.

(105)In general, a combination of vertical restraints aggravates their individual negative effects. However, certain combinations of vertical restraints are less anti-competitive than their use in isolation. For instance, in an exclusive distribution system, the distributor may be tempted to increase the price of the products as intra-brand competition has been reduced. The use of quantity forcing or the setting of a maximum resale price may limit such price increases. Possible negative effects of vertical restraints are reinforced when several suppliers and their buyers organise their trade in a similar way, leading to so-called cumulative effects.

1.2. Positive effects of vertical restraints

(106)It is important to recognise that vertical restraints may have positive effects by, in particular, promoting non-price competition and improved quality of services. When a company has no market power, it can only try to increase its profits by optimising its manufacturing or distribution processes. In a number of situations vertical restraints may be helpful in this respect since the usual arm's length dealings between supplier and buyer, determining only price and quantity of a certain transaction, can lead to a sub-optimal level of investments and sales.

(107)While trying to give a fair overview of the various justifications for vertical restraints, these Guidelines do not claim to be complete or exhaustive. The following reasons may justify the application of certain vertical restraints:

(a)To solve a 'free-rider' problem. One distributor may free-ride on the promotion efforts of another distributor. That type of problem is most common at the wholesale and retail level. Exclusive distribution or similar restrictions may be helpful in avoiding such free-riding. Free-riding can also occur between suppliers, for instance where one invests in promotion at the buyer's premises, in general at the retail level, that may also attract customers for its competitors. Non-compete type restraints can help to overcome free-riding [(41)].

For there to be a problem, there needs to be a real free-rider issue. Free-riding between buyers can only occur on pre-sales services and other promotional activities, but not on after-sales services for which the distributor can charge its customers individually. The product will usually need to be relatively new or technically complex or the reputation of the product must be a major determinant of its demand, as the customer may otherwise very well know what it wants, based on past purchases. And the product must be of a reasonably high value as it is otherwise not attractive for a customer to go to one shop for information and to another to buy. Lastly, it must not be practical for the supplier to impose on all buyers, by contract, effective promotion or service requirements.

Free-riding between suppliers is also restricted to specific situations, namely to cases where the promotion takes place at the buyer's premises and is generic, not brand specific.

(b)To 'open up or enter new markets'. Where a manufacturer wants to enter a new geographic market, for instance by exporting to another country for the first time, this may involve special 'first time investments' by the distributor to establish the brand on the market. In order to persuade a local distributor to make these investments, it may be necessary to provide territorial protection to the distributor so that it can recoup these investments by temporarily charging a higher price. Distributors based in other markets should then be

restrained for a limited period from selling on the new market (see also paragraph (61) in Section III.4). This is a special case of the free-rider problem described under point (a).

(c)The 'certification free-rider issue'. In some sectors, certain retailers have a reputation for stocking only 'quality' products. In such a case, selling through those retailers may be vital for the introduction of a new product. If the manufacturer cannot initially limit its sales to the premium stores, it runs the risk of being de-listed and the product introduction may fail. There may, therefore, be a reason for allowing for a limited duration a restriction such as exclusive distribution or selective distribution. It must be enough to guarantee introduction of the new product but not so long as to hinder large-scale dissemination. Such benefits are more likely with 'experience' goods or complex goods that represent a relatively large purchase for the final consumer.

(d)The so-called 'hold-up problem'. Sometimes there are client-specific investments to be made by either the supplier or the buyer, such as in special equipment or training. For instance, a component manufacturer that has to build new machines and tools in order to satisfy a particular requirement of one of its customers. The investor may not commit the necessary investments before particular supply arrangements are fixed.

However, as in the other free-riding examples, there are a number of conditions that have to be met before the risk of under-investment is real or significant. Firstly, the investment must be relationship-specific. An investment made by the supplier is considered to be relationship-specific when, after termination of the contract, it cannot be used by the supplier to supply other customers and can only be sold at a significant loss. An investment made by the buyer is considered to be relationship-specific when, after termination of the contract, it cannot be used by the buyer to purchase and/or use products supplied by other suppliers and can only be sold at a significant loss. An investment is thus relationship-specific because it can only, for instance, be used to produce a brand-specific component or to store a particular brand and thus cannot be used profitably to produce or resell alternatives. Secondly, it must be a long-term investment that is not recouped in the short run. And thirdly, the investment must be asymmetric, that is, one party to the contract invests more than the other party. Where these conditions are met, there is usually a good reason to have a vertical restraint for the duration it takes to depreciate the investment. The appropriate vertical restraint will be of the non-compete type or quantity-forcing type when the investment is made by the supplier and of the exclusive distribution, exclusive customer allocation or exclusive supply type when the investment is made by the buyer.

(e)The 'specific hold-up problem that may arise in the case of transfer of substantial know-how'. The know-how, once provided, cannot be taken back and the provider of the know-how may not want it to be used for or by its competitors. In as far as the know-how was not readily available to the buyer, is substantial and indispensable for the operation of the agreement, such a transfer may justify a non-compete type of restriction, which would normally fall outside Article 101(1).

(f)The 'vertical externality issue'. A retailer may not gain all the benefits of its action taken to improve sales; some may go to the manufacturer. For every extra unit a retailer sells by lowering its resale price or by increasing its sales effort, the manufacturer benefits if its wholesale price exceeds its marginal production costs. Thus, there may be a positive externality bestowed on the manufacturer by such retailer's actions and from the manufacturer's perspective the retailer may be pricing too high and/or making too little sales efforts. The negative externality of too high pricing by the retailer is sometimes called the "double marginalisation problem" and it can be avoided by imposing a maximum resale price on the retailer. To increase the retailer's sales efforts selective distribution, exclusive distribution or similar restrictions may be helpful [(42)].

(g)'Economies of scale in distribution'. In order to have scale economies exploited and thereby see a lower retail price for its product, the manufacturer may want to concentrate the resale of its products on a limited number of distributors. To do so, it could use exclusive distribution, quantity forcing in the form of a minimum purchasing requirement, selective distribution containing such a requirement or exclusive sourcing.

(h)'Capital market imperfections'. The usual providers of capital (banks, equity markets) may provide capital sub-optimally when they have imperfect information on the quality of the borrower or there is an inadequate basis to secure the loan. The buyer or supplier may have better information and be able, through an exclusive relationship, to obtain extra security for its investment. Where the supplier provides the loan to the buyer, this may lead to non-compete or quantity forcing on the buyer. Where the buyer provides the loan to the supplier, this may be the reason for having exclusive supply or quantity forcing on the supplier.

(i)'Uniformity and quality standardisation'. A vertical restraint may help to create a brand image by imposing a certain measure of uniformity and quality standardisation on the distributors, thereby increasing the attractiveness of the product to the final consumer and increasing its sales. This can for instance be found in selective distribution and franchising.

(108)The nine situations listed in paragraph (107) make clear that under certain conditions, vertical agreements are likely to help realise efficiencies and the development of new markets and that

this may offset possible negative effects. The case is in general strongest for vertical restraints of a limited duration which help the introduction of new complex products or protect relationship-specific investments. A vertical restraint is sometimes necessary for as long as the supplier sells its product to the buyer (see in particular the situations described in paragraph (107)(a), (e), (f), (g) and (i)).

(109)A large measure of substitutability exists between the different vertical restraints. As a result, the same inefficiency problem can be solved by different vertical restraints. For instance, economies of scale in distribution may possibly be achieved by using exclusive distribution, selective distribution, quantity forcing or exclusive sourcing. However, the negative effects on competition may differ between the various vertical restraints, which plays a role when indispensability is discussed under Article 101(3).

1.3. Methodology of analysis

(110)The assessment of a vertical restraint generally involves the following four steps ([43]):

(a)First, the undertakings involved need to establish the market shares of the supplier and the buyer on the market where they respectively sell and purchase the contract products.

(b)If the relevant market share of the supplier and the buyer each do not exceed the 30 % threshold, the vertical agreement is covered by the Block Exemption Regulation, subject to the hardcore restrictions and excluded restrictions set out in that Regulation.

(c)If the relevant market share is above the 30 % threshold for supplier and/or buyer, it is necessary to assess whether the vertical agreement falls within Article 101(1).

(d)If the vertical agreement falls within Article 101(1), it is necessary to examine whether it fulfils the conditions for exemption under Article 101(3).

1.3.1. Relevant factors for the assessment under Article 101(1)

(111)In assessing cases above the market share threshold of 30 %, the Commission will undertake a full competition analysis. The following factors are particularly relevant to establish whether a vertical agreement brings about an appreciable restriction of competition under Article 101(1):

(a) nature of the agreement;

(b) market position of the parties;

(c) market position of competitors;

(d) market position of buyers of the contract products;

(e) entry barriers;

(f) maturity of the market;

(g) level of trade;

(h) nature of the product;

(i) other factors.

(112)The importance of individual factors may vary from case to case and depends on all other factors. For instance, a high market share of the parties is usually a good indicator of market power, but in the case of low entry barriers it may not be indicative of market power. It is therefore not possible to provide firm rules on the importance of the individual factors.

(113)Vertical agreements can take many shapes and forms. It is therefore important to analyse the nature of the agreement in terms of the restraints that it contains, the duration of those restraints and the percentage of total sales on the market affected by those restraints. It may be necessary to go beyond the express terms of the agreement. The existence of implicit restraints may be derived from the way in which the agreement is implemented by the parties and the incentives that they face.

(114)The market position of the parties provides an indication of the degree of market power, if any, possessed by the supplier, the buyer or both. The higher their market share, the greater their market power is likely to be. This is particularly so where the market share reflects cost advantages or other competitive advantages vis-à-vis competitors. Such competitive advantages may, for instance, result from being a first mover on the market (having the best site, etc.), from holding essential patents or having superior technology, from being the brand leader or having a superior portfolio.

(115)Such indicators, namely market share and possible competitive advantages, are used to assess the market position of competitors. The stronger the competitors are and the greater their number, the less risk there is that the parties will be able to individually exercise market power and foreclose the market or soften competition. It is also relevant to consider whether there are effective and timely counterstrategies that competitors would be likely to deploy. However, if the number of competitors becomes rather small and their market position (size, costs, R&D potential, etc.) is rather similar, such a market structure may increase the risk of collusion. Fluctuating or rapidly changing market shares are in general an indication of intense competition.

(116)The market position of the parties' customers provides an indication of whether or not one or more of those customers possess buyer power. The first indicator of buyer power is the market share of the customer on the purchase market. That share reflects the importance of its

demand for possible suppliers. Other indicators focus on the position of the customer on its resale market, including characteristics such as a wide geographic spread of its outlets, own brands including private labels and its brand image amongst final consumers. In some circumstances, buyer power may prevent the parties from exercising market power and thereby solve a competition problem that would otherwise have existed. This is particularly so when strong customers have the capacity and incentive to bring new sources of supply on to the market in the case of a small but permanent increase in relative prices. Where strong customers merely extract favourable terms for themselves or simply pass on any price increase to their customers, their position does not prevent the parties from exercising market power.

(117)Entry barriers are measured by the extent to which incumbent companies can increase their price above the competitive level without attracting new entry. In the absence of entry barriers, easy and quick entry would render price increases unprofitable. When effective entry, preventing or eroding the exercise of market power, is likely to occur within one or two years, entry barriers can, as a general rule, be said to be low. Entry barriers may result from a wide variety of factors such as economies of scale and scope, government regulations, especially where they establish exclusive rights, state aid, import tariffs, intellectual property rights, ownership of resources where the supply is limited due to for instance natural limitations ([44]), essential facilities, a first mover advantage and brand loyalty of consumers created by strong advertising over a period of time. Vertical restraints and vertical integration may also work as an entry barrier by making access more difficult and foreclosing (potential) competitors. Entry barriers may be present at only the supplier or buyer level or at both levels. The question whether certain of those factors should be described as entry barriers depends particularly on whether they entail sunk costs. Sunk costs are those costs that have to be incurred to enter or be active on a market but that are lost when the market is exited. Advertising costs to build consumer loyalty are normally sunk costs, unless an exiting firm could either sell its brand name or use it somewhere else without a loss. The more costs are sunk, the more potential entrants have to weigh the risks of entering the market and the more credibly incumbents can threaten that they will match new competition, as sunk costs make it costly for incumbents to leave the market. If, for instance, distributors are tied to a manufacturer via a non-compete obligation, the foreclosing effect will be more significant if setting up its own distributors will impose sunk costs on the potential entrant. In general, entry requires sunk costs, sometimes minor and sometimes major. Therefore, actual competition is in general more effective and will weigh more heavily in the assessment of a case than potential competition.

(118)A mature market is a market that has existed for some time, where the technology used is well known and widespread and not changing very much, where there are no major brand

innovations and in which demand is relatively stable or declining. In such a market, negative effects are more likely than in more dynamic markets.

(119)The level of trade is linked to the distinction between intermediate and final goods and services. Intermediate goods and services are sold to undertakings for use as an input to produce other goods or services and are generally not recognisable in the final goods or services. The buyers of intermediate products are usually well-informed customers, able to assess quality and therefore less reliant on brand and image. Final goods are, directly or indirectly, sold to final consumers that often rely more on brand and image. As distributors have to respond to the demand of final consumers, competition may suffer more when distributors are foreclosed from selling one or a number of brands than when buyers of intermediate products are prevented from buying competing products from certain sources of supply.

(120)The nature of the product plays a role in particular for final products in assessing both the likely negative and the likely positive effects. When assessing the likely negative effects, it is important whether the products on the market are more homogeneous or heterogeneous, whether the product is expensive, taking up a large part of the consumer's budget, or is inexpensive and whether the product is a one-off purchase or repeatedly purchased. In general, when the product is more heterogeneous, less expensive and resembles more a one-off purchase, vertical restraints are more likely to have negative effects.

(121)In the assessment of particular restraints other factors may have to be taken into account. Among these factors can be the cumulative effect, that is, the coverage of the market by similar agreements of others, whether the agreement is 'imposed' (mainly one party is subject to the restrictions or obligations) or 'agreed' (both parties accept restrictions or obligations), the regulatory environment and behaviour that may indicate or facilitate collusion like price leadership, pre-announced price changes and discussions on the 'right' price, price rigidity in response to excess capacity, price discrimination and past collusive behaviour.

1.3.2. Relevant factors for the assessment under Article 101(3)

(122)Restrictive vertical agreements may also produce pro-competitive effects in the form of efficiencies, which may outweigh their anti-competitive effects. Such an assessment takes place within the framework of Article 101(3), which contains an exception from the prohibition rule of Article 101(1). For that exception to be applicable, the vertical agreement must produce objective economic benefits, the restrictions on competition must be indispensable to attain the efficiencies, consumers must receive a fair share of the efficiency gains, and the agreement must not afford the parties the possibility of eliminating competition in respect of a substantial part of the products concerned [(45)].

(123)The assessment of restrictive agreements under Article 101(3) is made within the actual context in which they occur ([46]) and on the basis of the facts existing at any given point in time. The assessment is sensitive to material changes in the facts. The exception rule of Article 101(3) applies as long as the four conditions are fulfilled and ceases to apply when that is no longer the case ([47]). When applying Article 101(3) in accordance with these principles it is necessary to take into account the investments made by any of the parties and the time needed and the restraints required to commit and recoup an efficiency enhancing investment.

(124)The first condition of Article 101(3) requires an assessment of what are the objective benefits in terms of efficiencies produced by the agreement. In this respect, vertical agreements often have the potential to help realise efficiencies, as explained in section 1.2, by improving the way in which the parties conduct their complementary activities.

(125)In the application of the indispensability test contained in Article 101(3), the Commission will in particular examine whether individual restrictions make it possible to perform the production, purchase and/or (re)sale of the contract products more efficiently than would have been the case in the absence of the restriction concerned. In making such an assessment, the market conditions and the realities facing the parties must be taken into account. Undertakings invoking the benefit of Article 101(3) are not required to consider hypothetical and theoretical alternatives. They must, however, explain and demonstrate why seemingly realistic and significantly less restrictive alternatives would be significantly less efficient. If the application of what appears to be a commercially realistic and less restrictive alternative would lead to a significant loss of efficiencies, the restriction in question is treated as indispensable.

(126)The condition that consumers must receive a fair share of the benefits implies that consumers of the products purchased and/or (re)sold under the vertical agreement must at least be compensated for the negative effects of the agreement. ([48]) In other words, the efficiency gains must fully off-set the likely negative impact on prices, output and other relevant factors caused by the agreement.

(127)The last condition of Article 101(3), according to which the agreement must not afford the parties the possibility of eliminating competition in respect of a substantial part of the products concerned, presupposes an analysis of remaining competitive pressures on the market and the impact of the agreement on such sources of competition. In the application of the last condition of Article 101(3), the relationship between Article 101(3) and Article 102 must be taken into account. According to settled case law, the application of Article 101(3) cannot prevent the application of Article 102 ([49]). Moreover, since Articles 101 and 102 both pursue the aim of maintaining effective competition on the market, consistency requires that Article 101(3) be

interpreted as precluding any application of the exception rule to restrictive agreements that constitute an abuse of a dominant position ([50]). The vertical agreement may not eliminate effective competition, by removing all or most existing sources of actual or potential competition. Rivalry between undertakings is an essential driver of economic efficiency, including dynamic efficiencies in the form of innovation. In its absence, the dominant undertaking will lack adequate incentives to continue to create and pass on efficiency gains. Where there is no residual competition and no foreseeable threat of entry, the protection of rivalry and the competitive process outweighs possible efficiency gains. A restrictive agreement which maintains, creates or strengthens a market position approaching that of a monopoly can normally not be justified on the grounds that it also creates efficiency gains.

2. Analysis of specific vertical restraints

(128)The most common vertical restraints and combinations of vertical restraints are analysed below following the framework of analysis developed in paragraphs 96 to 127. There are other restraints and combinations for which no direct guidance is provided here. They will however be treated according to the same principles and with the same emphasis on the effect on the market.

[…]

[IN EXTRACT BELOW ARE JUST TWO EXAMPLES OF EFFECTS-BASED ANALYSES GUIDANCE RELATED TO ASSESSING A PARTICULAR TYPE OF VERTICAL RESTRAINTS – I.E BUSINESS PRACTICES OF A PARTICULAR TYPE AND FOR BROADLY COMMON BUSINESS PURPOSES - IN THESE GUIDELINES]

2.8. Category Management Agreements

(209)Category management agreements are agreements by which, within a distribution agreement, the distributor entrusts the supplier (the 'category captain') with the marketing of a category of products including in general not only the supplier's products, but also the products of its competitors. The category captain may thus have an influence on for instance the product placement and product promotion in the shop and product selection for the shop. Category management agreements are exempted under the Block Exemption Regulation when both the supplier's and buyer's market share does not exceed 30 %. The remainder of this section provides guidance for the assessment of category management agreements in individual cases above the market share threshold.

(210)While in most cases category management agreements will not be problematic, they may sometimes distort competition between suppliers, and finally result in anticompetitive foreclosure of other suppliers, where the category captain is able, due to its influence over

the marketing decisions of the distributor, to limit or disadvantage the distribution of products of competing suppliers. While in most cases the distributor may not have an interest in limiting its choice of products, when the distributor also sells competing products under its own brand (private labels), the distributor may also have incentives to exclude certain suppliers, in particular intermediate range products. The assessment of such upstream foreclosure effect is made by analogy to the assessment of single branding obligations (in particular paragraphs (132) to (141)) by addressing issues like the market coverage of these agreements, the market position of competing suppliers and the possible cumulative use of such agreements.

(211)In addition, category management agreements may facilitate collusion between distributors when the same supplier serves as a category captain for all or most of the competing distributors on a market and provides these distributors with a common point of reference for their marketing decisions.

(212)Category management may also facilitate collusion between suppliers through increased opportunities to exchange via retailers sensitive market information, such as for instance information related to future pricing, promotional plans or advertising campaigns (58).

(213)However, the use of category management agreements may also lead to efficiencies. Category management agreements may allow distributors to have access to the supplier's marketing expertise for a certain group of products and to achieve economies of scale as they ensure that the optimal quantity of products is presented timely and directly on the shelves. As category management is based on customers' habits, category management agreements may lead to higher customer satisfaction as they help to better meet demand expectations. In general, the higher the inter-brand competition and the lower consumers' switching costs, the greater the economic benefits achieved through category management.

2.9 Tying

(214) Tying refers to situations where customers that purchase one product (the tying product) are required also to purchase another distinct product (the tied product) from the same supplier or someone designated by the latter. Tying may constitute an abuse within the meaning of Article 102 (59). Tying may also constitute a vertical restraint falling under Article 101 where it results in a single branding type of obligation (see paragraphs (129) to (150)) for the tied product. Only the latter situation is dealt with in these Guidelines.

(215)Whether products will be considered as distinct depends on customer demand. Two products are distinct where, in the absence of the tying, a substantial number of customers would

purchase or would have purchased the tying product without also buying the tied product from the same supplier, thereby allowing stand-alone production for both the tying and the tied product (60). Evidence that two products are distinct could include direct evidence that, when given a choice, customers purchase the tying and the tied products separately from different sources of supply, or indirect evidence, such as the presence on the market of undertakings specialised in the manufacture or sale of the tied product without the tying product (61), or evidence indicating that undertakings with little market power, particularly on competitive markets, tend not to tie or not to bundle such products. For instance, since customers want to buy shoes with laces and it is not practicable for distributors to lace new shoes with the laces of their choice, it has become commercial usage for shoe manufacturers to supply shoes with laces. Therefore, the sale of shoes with laces is not a tying practice.

(216)Tying may lead to anticompetitive foreclosure effects on the tied market, the tying market, or both at the same time. The foreclosure effect depends on the tied percentage of total sales on the market of the tied product. On the question of what can be considered appreciable foreclosure under Article 101(1), the analysis for single branding can be applied. Tying means that there is at least a form of quantity-forcing on the buyer in respect of the tied product. Where in addition a non-compete obligation is agreed in respect of the tied product, this increases the possible foreclosure effect on the market of the tied product. The tying may lead to less competition for customers interested in buying the tied product, but not the tying product. If there is not a sufficient number of customers that will buy the tied product alone to sustain competitors of the supplier on the tied market, the tying can lead to those customers facing higher prices. If the tied product is an important complementary product for customers of the tying product, a reduction of alternative suppliers of the tied product and hence a reduced availability of that product can make entry onto the tying market alone more difficult.

(217)Tying may also directly lead to prices that are above the competitive level, especially in three situations. Firstly, if the tying and the tied product can be used in variable proportions as inputs to a production process, customers may react to an increase in price for the tying product by increasing their demand for the tied product while decreasing their demand for the tying product. By tying the two products the supplier may seek to avoid this substitution and as a result be able to raise its prices. Secondly, when the tying allows price discrimination according to the use the customer makes of the tying product, for example the tying of ink cartridges to the sale of photocopying machines (metering). Thirdly, when in the case of long-term contracts or in the case of after-markets with original equipment with a long replacement time, it becomes difficult for the customers to calculate the consequences of the tying.

(218)Tying is exempted under the Block Exemption Regulation when the market share of the supplier, on both the market of the tied product and the market of the tying product, and the market share of the buyer, on the relevant upstream markets, do not exceed 30 %. It may be combined with other vertical restraints, which are not hardcore restrictions under that Regulation, such as non-compete obligations or quantity forcing in respect of the tying product, or exclusive sourcing. The remainder of this section provides guidance for the assessment of tying in individual cases above the market share threshold.

219) The market position of the supplier on the market of the tying product is obviously of central importance to assess possible anti-competitive effects. In general, this type of agreement is imposed by the supplier. The importance of the supplier on the market of the tying product is the main reason why a buyer may find it difficult to refuse a tying obligation.

(220)The market position of the supplier's competitors on the market of the tying product is important in assessing the supplier's market power. As long as its competitors are sufficiently numerous and strong, no anti-competitive effects can be expected, as buyers have sufficient alternatives to purchase the tying product without the tied product, unless other suppliers are applying similar tying. In addition, entry barriers on the market of the tying product are relevant to establish the market position of the supplier. When tying is combined with a non-compete obligation in respect of the tying product, this considerably strengthens the position of the supplier.

(221)Buying power is relevant, as important buyers will not easily be forced to accept tying without obtaining at least part of the possible efficiencies. Tying not based on efficiency is therefore mainly a risk where buyers do not have significant buying power.

(222)Where appreciable anti-competitive effects are established, the question whether the conditions of Article 101(3) are fulfilled arises. Tying obligations may help to produce efficiencies arising from joint production or joint distribution. Where the tied product is not produced by the supplier, an efficiency may also arise from the supplier buying large quantities of the tied product. For tying to fulfil the conditions of Article 101(3), it must, however, be shown that at least part of these cost reductions are passed on to the consumer, which is normally not the case when the retailer is able to obtain, on a regular basis, supplies of the same or equivalent products on the same or better conditions than those offered by the supplier which applies the tying practice. Another efficiency may exist where tying helps to ensure a certain uniformity and quality standardisation (see paragraph (107)(i)). However, it needs to be demonstrated that the positive effects cannot be realised equally efficiently by

requiring the buyer to use or resell products satisfying minimum quality standards, without requiring the buyer to purchase these from the supplier or someone designated by the latter. The requirements concerning minimum quality standards would not normally fall within the scope of Article 101(1). Where the supplier of the tying product imposes on the buyer the suppliers from which the buyer must purchase the tied product, for instance because the formulation of minimum quality standards is not possible, this may also fall outside the scope of Article 101(1), especially where the supplier of the tying product does not derive a direct (financial) benefit from designating the suppliers of the tied product.

---

(1) With effect from 1 December 2009, Articles 81 and 82 of the EC Treaty have become Articles 101 and, 102, respectively, of the Treaty on the Functioning of the European Union ('TFEU'). The two sets of provisions are, in substance, identical. For the purposes of these Guidelines, references to Articles 101 and 102 of the TFEU should be understood as references to Articles 81 and 82, respectively, of the EC Treaty where appropriate. The TFEU also introduced certain changes in terminology, such as the replacement of 'Community' by 'Union' and 'common market' by 'internal market'. The terminology of the TFEU will be used throughout these Guidelines.

(2) These Guidelines replace the Commission Notice – Guidelines on Vertical Restraints, OJ C 291, 13.10.2000, p. 1.

(3) OJ L 102, 23.4.2010, p. 1.

(4) See inter alia *judgments of the Court of Justice in Joined Cases 56/64 and 58/64* Grundig-Consten v Commission *[1966] ECR 299; Case 56/65* Technique Minière *v* Maschinenbau Ulm *[1966] ECR 235; and judgment of the Court of First Instance in Case T-77/92* Parker Pen *v* Commission *[1994] ECR II-549.*

(5) See Communication from the Commission - Notice – Guidelines on the application of Article 81(3) of the Treaty, OJ C 101, 27.4.2004, p. 97 for the Commission's general methodology and interpretation of the conditions for applying Article 101(1) and in particular Article 101(3).

(6) OJ C 368, 22.12.2001, p. 13.

(7) For agreements between competing undertakings the de minimis *market share threshold is 10 % for their collective market share on each affected relevant market.*

(8) See judgment of the Court of First Instance in Case T-7/93 Langnese-Iglo *v* Commission *[1995] ECR II-1533, paragraph 98.*

(9) See judgments of the Court of Justice in Case 5/69 Völk *v* Vervaecke *[1969] ECR 295; Case 1/71* Cadillon *v* Höss *[1971] ECR 351 and Case C-306/96* Javico *v* Yves Saint Laurent *[1998] ECR I-1983, paragraphs 16 and 17.*

(10) OJ L 124, 20.5.2003, p. 36.

[…]

(33) Judgment of the Court of Justice of 28 February 1991 in Case C-234/89, Stergios Delimitis *v* Henninger Bräu AG *[1991] ECR I-935.*

(38) OJ L 1, 4.1.2003, p. 1.

(39) See Section II.1.

(40) By collusion is meant both explicit collusion and tacit collusion (conscious parallel behaviour).

(41) Whether consumers actually benefit overall from extra promotional efforts depends on whether the extra promotion informs and convinces and thus benefits many new customers or mainly reaches customers who already know what they want to buy and for whom the extra promotion only or mainly implies a price increase.

(42) See however the previous footnote.

(43) These steps are not intended to present a legal reasoning that the Commission should follow in this order to take a decision.

(44) See Commission Decision 97/26/EC (Case No IV/M.619 — Gencor/Lonrho*), OJ L 11, 14.1.1997, p. 30.*

(45) See Communication from the Commission - Notice – Guidelines on the application of Article 81(3) of the Treaty, OJ C 101, 27.4.2004, p. 97.

(46) See Judgment of the Court of Justice in Joined Cases 25/84 and 26/84 Ford *[1985] ECR 2725.*

(47) See in this respect for example Commission Decision 1999/242/EC (Case No IV/36.237 – TPS*), OJ L 90, 2.4.1999, p. 6. Similarly, the prohibition of Article 101(1) also only applies as long as the agreement has a restrictive object or restrictive effects.*

(48) See paragraph 85 of Communication from the Commission - Notice – Guidelines on the application of Article 81(3) of the Treaty, OJ C 101, 27.4.2004, p. 97.

(49) See Judgment of the Court of Justice in Joined Cases C-395/96 P and C-396/96 P Compagnie Maritime Belge *[2000] ECR I-1365, paragraph 130. Similarly, the application of Article 101(3) does not prevent the application of the Treaty rules on the free movement of goods, services, persons and capital. These provisions are in certain circumstances applicable to agreements, decisions and*

concerted practices within the meaning of Article 101(1), see to that effect Judgment of the Court of Justice in Case C-309/99 Wouters *[2002] ECR I-1577, paragraph 120.*

([50]) See in this respect Judgment of the Court of First Instance in Case T-51/89 Tetra Pak (I) *[1990] ECR II-309. See also paragraph 106 of Communication from the Commission - Notice – Guidelines on the application of Article 81(3) of the Treaty, OJ C 101, 27.4.2004, p. 97.*

[…]

(58) Direct information exchange between competitors is not covered by the Block Exemption Regulation, see Article 2(4) of that Regulation and paragraphs 27-28 of these Guidelines.

(59) Judgment of the Court of Justice in Case C-333/94 P Tetrapak v Commission [1996] ECR I-5951, paragraph 37. See also Communication from the Commission – Guidance on the Commission's enforcement priorities in applying Article 82 of the EC Treaty to abusive conduct by dominant undertakings, OJ C 45, 24.2.2009, p. 7.

(60) Judgment of the Court of First Instance in Case T-201/04 Microsoft v Commission [2007] ECR II-3601, paragraphs 917, 921 and 922.

(61) Judgment of the Court of First Instance in Case T-30/89 Hilti v Commission [1991] ECR II-1439, paragraph 67.

Bibliography

Legislation

Charter of Fundamental Rights of the European Union (2009), OJ 2010 C83/391

Commission Decision (EC) No 2000/520 of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles, OJ 2000 L215/7

Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield, OJ 2016 L207/1

Commission Regulation (EC) No 800/2008 of 6 August 2008 declaring certain categories of aid compatible with the common market in application of Articles 87 and 88 of the Treaty, OJ 2008 L214/3

Commission Regulation (EU) No 330/2010 of 20 April 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of vertical agreements and concerted practices, OJ 2010 L102/1

Commission Regulation (EU) No 651/2014 of 17 June 2014 declaring certain categories of aid compatible with the internal market in application of Articles 107 and 108 of the Treaty, OJ 2014 L187/1

Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, (1981), CETS No. 108

Council Regulation No 17 (EEC): First Regulation implementing Articles 85 and 86 of the Treaty [at present Articles 101 and 102 TFEU]

Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty, OJ 2003 L1/1 (the Modernisation Regulation)

Data Protection Act (Datenschutzgesetz) 2000 (Austria)

Data Protection Act 1988 (Ireland, rev. 2003)

Data Protection Act 1998 (UK)

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

Federal Trade Commission Act of 1914 (US)

Freedom of Information Act 2000 (UK)

Freedom of Information (Scotland) Act 2002

Health insurance portability and accountability Act of 1996 (US)

Personal Information Protection and Electronic Documents Act 2000 (Canada, rev. 2015)

Privacy Act 1988 (Australia)

Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC

Regulation (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 Directive 95/46/EC

The European Convention on Human Rights (1950)

The International Covenant on Civil and Political Rights (1966)

Treaty establishing the European Economic Community (the Treaty of Rome, 1957)

Treaty on European Union (the Maastricht Treaty, 1992)

Treaty of Amsterdam amending the Treaty on European Union (1999)

Treaty on the Functioning of the European Union (the Treaty of Lisbon, 2007)

The Universal Declaration of Human Rights (1948)

Draft Legislation

Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM/2012/011 final

European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), P7_TA(2014)0212

Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) — Preparation of a general approach, Council document 9565/15 (11 June 2015)

Regulator and Other Policy Publications

Internationally

Australian Law Reform Commission, For Your Information: Australian Privacy Law and Practice; Report. Law Reform Commission., vol.1, (2008) **["Australian LRC, 2008, #1"]**

Australian Law Reform Commission, What is Personal Information? Report 108, Volume 1 (2008), available at: http://www.alrc.gov.au/sites/default/files/pdfs/108_vol1.pdf [Accessed 1 August 2017] **["Australian LRC, 2008, #2"]**

OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. (1980, rev.2013)

US Office for Civil Rights, 2002. Standards for privacy of individually identifiable health information. Final rule. Federal Register, 67(157), p.53181, §164.514 (a) ('the HIPPA Privacy Rule')

US Office for Civil Rights, 26 November 2012. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, available at: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf, [Accessed 1 August 2017]

US Federal Trade Commission, Privacy Report: Protecting consumer privacy in an era of rapid change: Recommendations for businesses and policymakers. (2012)

EU

Article 29 Working Party, Working Document - Privacy on the Internet – An integrated EU Approach to Online Data Protection, 21 November 2000, WP37

Bibliography

Article 29 Working Party, Working document on data protection issues related to RFID technology, 19 January 2005, WP105

Article 29 Working Party, Opinion 04/2007 on the concept of personal data, 20 June 2007, WP136

Article 29 Working Party, Opinion 1/2008 on data protection issues relating to search engines, 4 April 2008, WP148

Article 29 Working Party, Opinion 3/2010 on the principle of accountability, 13 July 2010, WP173

Article 29 Working Party, Opinion 13/2011 on geo-location services on smart devices, 16 May 2011, WP185

Article 29 Working Party, 2011, Privacy and Data Protection Impact Assessment Framework for RFID Applications, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2011/wp180_annex_en.pdf [Accessed 1 August DATE 2017]

Article 29 Working Party, Opinion 1/2012 on the data protection reform proposals, 23 March 2012, WP191

Article 29 Working Party, Opinion 03/2012 on developments in biometric technologies, 27 April 2012, WP193

Article 29 Working Party, Advice paper of 13 May 2013 on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation (2013), available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130513_advice-paper-on-profiling_en.pdf, [Accessed 1 August 2017]

Article 29 Working Party, Statement of 27 February 2013 on current discussions regarding the data protection reform package, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130227_statement_dp_reform_package_en.pdf, [Accessed: 1 August 2017]

Article 29 Working Party, Opinion 3/2013 on Purpose Limitation, 2 April 2013, WP203

Article 29 Working Party, Opinion 6/2013 on open data and public sector information ('PSI') reuse, 5 June 2013, WP207

Article 29 Working Party, Opinion 7/2013 on the Data Protection Impact Assessment Template for Smart Grid and Smart Metering Systems ('DPIA Template') prepared by Expert Group 2 of the Commission's Smart Grid Task Force, 4 December 2013, WP209

Article 29 Working Party, Opinion 6/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 9 April 2014, WP217

Article 29 Working Party, Opinion 05/2014 on anonymisation techniques, 10 April 2014, WP216

Article 29 Working Party, Statement on the role of a risk-based approach in data protection legal frameworks, 30 May 2014, WP218

Article 29 Working Party, Guidelines for identifying a controller or processor's lead supervisory authority, 31 December 2016, WP244

Article 29 Working Party, Draft 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, 4 April 2017, WP248

Article 29 Working Party, 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, 4 October 2017, WP248 rev.01

Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 3 October 2017, WP251

European Commission, Analysis and impact study on the implementation of Directive EC 95/46 in Member States, accompanying the First Report on the implementation of the Data Protection Directive (95/46/EC), May 2003, COM (2003) 265 final

European Commission, A comprehensive approach on personal data protection in the European Union, November 2010, (Communication) COM (2010) 609 final

European Commission Notice, 2004, Guidelines on the effect on trade concept contained in Articles 81 and 82 of the Treaty [ex Articles 101 and 102 TFEU], OJ 2004 C101/81

European Commission Notice, 2010, Guidelines on Vertical Restraints, OJ 2010 C130/1

European Commission, 2014, Notice On agreements of minor importance which do not appreciably restrict competition under Article 101(1) of the Treaty on the Functioning of the European Union, OJ 2014 C291/1

European Commission staff working paper Impact Assessment /* SEC/2012/0072 final - Accompanying the document Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) {COM(2012) 10 final}, available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012SC0072, [Accessed 1 August 2017]

European Data Protection Supervisor. (2014). Preliminary Opinion of the European Data Protection Supervisor – Privacy and competitiveness in the age of bid data: The interplay between data protection, competition law and consumer protection interaction in the Digital Economy, available at: https://edps.europa.eu/sites/edp/files/publication/14-03-26_competitition_law_big_data_en.pdf [Accessed 1 August 2017]

European Data Protection Supervisor. (2015). Recommendations on the EU's options for data protection reform, available at: https://edps.europa.eu/sites/edp/files/publication/15-07-27_gdpr_summary_en_0.pdf [Accessed 1 August 2017]

European Parliament Committee on Civil Liberties, Justice and Home Affairs. Draft report on the proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)), PR\922387EN.doc, (17 December 2012)

UK

Information Commissioner's response to the Commission Consultation to the Legal Framework for the Fundamental Right to Protection of Personal Data, 2009, available at: http://ec.europa.eu/justice/news/consulting_public/0003/contributions/public_authorities/ico_uk_en.pdf, [Accessed 1 August 2017]

Information Commissioner's response to the Ministry of Justice's call for evidence on the current data protection legislative framework, 2010, available at: http://www.statewatch.org/news/2010/oct/ul-ico-response-eu-dp-review.pdf [Accessed 1 August 2017]

Information Commissioner's response to 'A comprehensive approach on personal data protection in the European Union - A Communication from the European Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions on 4 November 2010', (2011), available at:

http://ec.europa.eu/justice/news/consulting_public/0006/contributions/public_authorities/ico_infocommoffice_en.pdf [1 August DATE 2017]

Information Commissioner's response to the House of Commons Science and Technology Committee inquiry on "The big data dilemma", 2015, available at: https://ico.org.uk/media/about-the-ico/consultation-responses/2015/1432826/ico-response-hc-science-and-technology-committee-consultation-on-big-data.pdf [Accessed 1 August 2017]

Information Commissioner's Office. Data Protection Act 1998 Legal Guidance. (2001), available at: http://www.valident.co.uk/wp-content/uploads/2012/01/data_protection_act_legal_guidance.pdf [1 August DATE 2017].

Information Commissioner's Office. Data Protection Technical Guidance: Determining what is Personal Data. (2007), available at: https://ico.org.uk/media/1554/determining-what-is-personal-data.pdf [Accessed 1 August 2017]

Information Commissioner's Office. Privacy Impact Assessment Handbook (2007, rev. 2009)

Information Commissioner's Office. Personal Information Code of Practice (2010), available at: https://ico.org.uk/media/for-organisations/documents/1591/personal_information_online_cop.pdf, [Accessed 1 August 2017]

Information Commissioner's Office. Initial analysis of the European Commission's proposals for a revised data protection legislative framework, 27 February 2012, available at http://www.ico.gov.uk/~/media/documents/library/Data_Protection/Research_and_reports/ico_initial_analysis_of_revised_eu_dp_legislative_proposals.ashx, [Accessed 1 August 2017]

Information Commissioner's Office. Anonymisation: Managing Data Protection Risk Code of Practice (2012), available at: https://ico.org.uk/media/1061/anonymisation-code.pdf, [Accessed 1 August 2017]

Information Commissioner's Office. Proposed new EU General Data Protection Regulation: Article-by-article analysis paper, V1.0 12 February 2013, available at: https://ico.org.uk/media/about-the-ico/documents/1042564/ico-proposed-dp-regulation-analysis-paper-20130212.pdf, [Accessed 1 August 2017]

Information Commissioner's Office. Written evidence (SMD0018) submitted to the House of Commons Science and Technology Committee in response to its inquiry into social media data and real time analytics (2014), available at: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-

and-technology-committee/social-media-data-and-real-time-analytics/written/8171.html [Accessed 1 August 2017]

Information Commissioner's Office. Big Data and Data Protection (2014), available at www.pdpjournals.com/88383.pdf [Accessed 1 August 2017] **["ICO, 2014, #1"; rev. 2017 as below]**

Information Commissioner's' Office. Conducting Privacy Impact Assessments Code of Practice (2014), available at: https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf [Accessed 1 August 2017] **["ICO, 2014, #2"]**

Information Commissioner's Office. Feedback request – profiling and automated decision-making (2017), available at: https://ico.org.uk/media/2013894/ico-feedback-request-profiling-and-automated-decision-making.pdf [Accessed 1 August DATE 2017] **["ICO, 2017, #1"]**

Information Commissioner's Office. Big Data, Artificial Intelligence, Machine Learning and Data Protection, available at: http://ico.org.uk/news/latest_news/2014/~/media/documents/library/Data_Protection/Practical_application/big-data-and-data-protection.pdf [Accessed 1 August 2017] **["ICO, 2017, #2"]**

RCUK (UK Research Councils)'s Response to the Ministry of Justice's call for evidence on the current data protection legislative framework (2010), available at: http://www.rcuk.ac.uk/documents/submissions/mojdataprotectionlegislative-oct2010-pdf/ [Accessed 1 August 2017]

Thesis-Author Journal Papers

Knight, A., 2011. Out with the Old, in with the New-Comparing 1992 US Horizontal Merger Guidelines with the 2010 US Horizontal Merger Guidelines. Southampton Student L. Rev., 1, p.80

Knight, A. and Saxby, S., 2014. Identity crisis: Global challenges of identity protection in a networked world. Computer Law & Security Review, 30(6), p.617

Knight, A., Pearce, H., and Saxby, S., 2014. 'Piercing the anonymity veil: re-identification risk and the UK transparency agenda', in Kierkegaard, Sylvia (ed.) Information Ethics and Security: Future of International World Time. Copenhagen, DK, International Association of IT Lawyers

Knight, A., and Stalla-Bourdillon, S., 2016. Anonymous data v. Personal data–A false debate: An EU perspective on anonymisation, pseudonymisation and personal data. Wis. Int'l LJ, p.284

Journal and Research Papers

Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., Sussman, G.J., and Weitzner, D.J., 2008. Information accountability. Communications of the ACM, 51(6), p.82

Aldhouse, F. (2014). Anonymisation of personal data–A missed opportunity for the European Commission. Computer Law & Security Review, 30(4), p.403

Ambrose, M.L., 2012. It's about time: privacy, information life cycles, and the right to be forgotten. *Stan. Tech. L. Rev.*, *16*, p.369

Balboni, P., Cooper, D., Imperiali, R. and Macenaite, M., 2013. Legitimate interest of the data controller New data protection paradigm: legitimacy grounded on appropriate protection. *International Data Privacy Law*, *3*(4), p.244

Beyleveld, D. and Townend, D.M., 2004. When is personal data rendered anonymous? Interpreting Recital 26 of Directive 95/46/EC. Medical law international, 6(2), p.73

Bergkamp, L. and Dhont, J., 2000. Data Protection in Europe and the Internet: An Analysis of the European Community's Privacy Legislation in the Context of the World Wide Web. EDI L. Rev., 7, p.71

Bing, J., 1972. Classification of personal information with respect to the sensitivity aspect. In Proceedings of the First International Oslo Symposium on Data Banks and Society

Booth, S., Jenkins, R., Moxon, D., Semmens, N., Spencer, C., Taylor, M. and Townend, D., 2004. What are 'Personal Data'? A study conducted for the UK Information Commissioner. Sheffield: The University of Sheffield, available at: http://www.frareg.com/news/documentazione/gestione/personal_data.pdf [Accessed 1 August 2017]

Bridges, P., 2015. EU and US privacy experts in search of transatlantic privacy solutions. Amsterdam/Cambridge, September, available at: https://privacybridges.mit.edu/sites/default/files/documents/PrivacyBridges-FINAL.pdf [Accessed 1 August 2017]

Burri, M. and Schär, R., 2016. The reform of the EU data protection framework: outlining key changes and assessing their fitness for a data-driven economy. Journal of Information Policy, 6(1), p.479

Bygrave, L.A., 2010. Privacy and data protection in an international perspective. Scandinavian studies in law, 56, p.165

Calo, M.R., 2011. The Boundaries of Privacy Harm. Indiana Law Journal, 86, p.1131

Castro, D. and Cavoukian, A., 2014. Big Data and innovation, setting the record straight: De-identification does Work. White Paper, Jun.

Cate, F.H., Kuner, C., Millard, C., Svantesson, D.J.B. and Lynskey, O., 2015. Risk management in data protection (editorial). International Data Privacy Law, 2015, Vol. 5, No. 2, p.95

Cate, F.H. and Mayer-Schönberger, V., 2013. Data Use and Impact Global Workshop. Center for Applied Cybersecurity Research, available at: https://iapp.org/resources/article/data-use-and-impact-global-workshop/ [Accessed 1 August 2017] **["Cate & Mayer-Schönberger, 2013, #2"]**

Costa-Cabral, F. and Lynskey, O., 2017. Family ties: the intersection between data protection and competition in EU Law. Common Market Law Review, 54(1), p.11

Creese, S., and Hodges, D., 2013, October. Breaking the Arc: Risk control for Big Data. In Big Data, 2013 IEEE International Conference on. IEEE, p.613.

Crompton, M., 2002. Under the Gaze, Privacy Identity and New Technology'. Paper presented at International Association of Lawyers 75th Anniversary Congress, Sydney, 28 October 2002

Cuijpers, C.M.C.K., 2004. The cookie controversy: a better solution with the new regulation? Regulation, p.21

De Hert, P., Kloza, D., Wright, D., Wadhwa, K., Hosein, G. and Davies, S., 2012. Recommendations for a privacy impact assessment framework for the European Union. Deliverable, PIAF project

De Montjoye, Y.A., Radaelli, L. and Singh, V.K., 2015. Unique in the shopping mall: On the reidentifiability of credit card metadata. Science, 347(6221), p.536

Dinda, S., 2016. Adaptation to Climate Change for Sustainable Development: A Survey. In: Information Resources Management Association (ed.), *Natural Resources Management: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, place: Hershey: IGI Global, p.334

Einav, L. and Levin, J., 2013. The Data Revolution and Economic Analysis (No. 19035). National Bureau of Economic Research, Inc.

El Emam, K., 2013. Guide to the de-identification of personal health information. CRC Press

El Emam, K. and Álvarez, C., 2014. A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques. International Data Privacy Law, p.ipu033

Elliot, M., Mackey, E., O'Hara, K. and Tudor, C., 2016. The Anonymisation Decision-Making Framework, available at: https://eprints.soton.ac.uk/399692/1/The-Anonymisation-Decision-making-Framework.pdf [Accessed 1 August 2017]

Felten, E.W. and Narayanan, A., 2014. No silver bullet: De-identification still doesn't work. White Paper

Floridi, L., 2014. Open data, data protection, and group privacy. Philosophy & Technology, 27(1), p.1

Floridi, L. and Mittelstadt, B.D., 2016. The ethics of big data: Current and foreseeable issues in biomedical contexts. Science and Engineering Ethics, 22(2), p.303

Fried, C., 1968. Privacy, Yale LJ, 77, p.475

Friedewald, M., Wright, D., Gutwirth, S. and Mordini, E., 2010. Privacy, data protection and emerging sciences and technologies: towards a common framework. Innovation–The European Journal of Social Science Research, 23(1), p.61

Future of Privacy Forum, Comments of the FPF to the FTC on Consumer Generated and Controlled Health Data, FTC Project No. P145401 (2014), Spring Privacy Series, available at: https://fpf.org/wp-content/uploads/FPF-Comments-to-FTC-on-CGHD.pdf, [Accessed 1 August 2017]

Gratton, E., 2013. If Personal Information is Privacy's Gatekeeper, Then Risk of Harm Is the Key: A Proposed Method for Determining What Counts as Personal Information. Alb. LJ Sci. & Tech., 24, p.105

Gellert, R. and Gutwirth, S., 2013. The legal construction of privacy and data protection. Computer Law & Security Review, 29(5), p.522

Hansen, M. and Pfitzmann, A., 2010. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, available at: https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf [Accessed 1 August 2017]

Hartzog, W., 2014. The Value of Modest Privacy Protections in a Hyper Social World. Colo. Tech. LJ, 12, p.333

Hartzog, W., and Rubinstein, I., 2016. Anonymization and risk. Washington Law Review, 91(2), p.703

Hartzog, W. and Stutzman, F., 2013. The case for online obscurity. Cal. L. Rev., 101, p.1 **["Hartzog & Stutzman, 2013, #1"]**

Hartzog, W. and Stutzman, F., 2013. Obscurity by design. Wash. L. Rev., 88, p.385 **["Hartzog & Stutzman, 2013, #2"]**

Hildebrandt, M. and Koop, B., 2007. FIDIS Deliverable D7. 9: A Vision of Ambient Law, available at http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp7-d7.9_A_Vision_of_Ambient_Law.pdf [Accessed 1 August 2017]

Hon, W.K., Kosta, E., Millard, C. and Stefanatou, D., 2014. Cloud accountability: The likely impact of the proposed EU data protection regulation. Queen Mary School of Law Legal Studies Research Paper No. 172/2014; Tilburg Law School Research Paper No. 07/2014. SSRN Repository, available at: https://ssrn.com/abstract=2405971 or http://dx.doi.org/10.2139/ssrn.2405971

Hon, W.K., Millard, C. and Walden, I., 2011. The problem of 'personal data' in cloud computing: what information is regulated?—the cloud of unknowing. Queen Mary School of Law Legal Studies Research Paper No. 75/2011. International Data Privacy Law, 1(4), p. 211

Hurley, D., Taking the Long Way Home: The Human Right of Privacy. In Rotenberg, M., Scott, J. and Horwitz, J. eds., 2015. Privacy in the modern age: The search for solutions. The New Press, p.70

Imperiali, R., 2012. The data protection compliance program. J. Int'l Com. L. & Tech., 7, p.285

Inness, J.C., 1996. Privacy, intimacy, and isolation. Oxford University Press on Demand, p.140

Korff, D., 2003. Study on Implementation of Data Protection Directive: Comparative Summary of National Laws. Human Rights Centre, University of Essex

Kuner, C., 2012. The European Commission's proposed data protection regulation: A copernican revolution in European data protection law, available at http://robertgrzeszczak.bio.wpia.uw.edu.pl/files/2012/12/Kuner_A-Copernican-Revolution-in-European-Data-Protection-Law.pdf [Accessed 1 August 2017]

Kupritz, V.W., 1998. Privacy in the work place: The impact of building design. Journal of Environmental Psychology, 18(4), p.341

Mantelero, A., 2014. Defining a new paradigm for data protection in the world of Big Data analytics. In: 2014 ASE BIGDATA/SOCIALCOM/CYBER SECURITY Conference, Stanford University, 27-31 May 2014, p.1

Mantelero, A., 2015. Data protection, e-ticketing, and intelligent systems for public transport. International Data Privacy Law, 5(4), p.309

Marx, G.T., Coming to Terms and Avoiding Information Techno-Fallacies. In Rotenberg, M., Scott, J. and Horwitz, J. eds., 2015. Privacy in the modern age: The search for solutions. The New Press, p.118

McCullagh, K., 2007. Data sensitivity: Proposals for resolving the conundrum. J. Int'l Com. L. & Tech, 2

McCullagh, K., 2009. Protecting 'privacy' through control of 'personal' data processing: A flawed approach. International Review of Law, Computers & Technology, 23(1-2)

Milaj, J., 2015. Invalidation of the data retention directive–extending the proportionality test. Computer Law & Security Review, 31(5), p.604

Nissenbaum, H., 2004. Privacy as contextual integrity. Wash. L. Rev., 79, p.119

Nissenbaum, H., 2011. A Contextual Approach to Privacy Online. Journal of the American Academy of Arts & Sciences, 140(4), p.32

Nissenbaum, H., 2014. Respect for Context as a Benchmark for Privacy Online: What it Is and Isn't. Cahier de prospective, p.19

Ohm, P., 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. (2010) 57. UCLA L Rev, 6, p.1701

Ohm, P., 2012. The underwhelming benefits of big data. U. Pa. L. Rev. PENNumbra, 161, p.339

Oswald, M., 2013. Something Bad Might Happen: Lawyers, anonymization and risk. XRDS: Crossroads, The ACM Magazine for Students, 20(1), p.22

Polonetsky, J. and Tene, O., Big Data for All: Privacy and User Control in the Age of Analytics' (2013). Northwestern Journal of Technology and Intellectual Property, 11, p.239

Polentsky, J. and Wolf, C., 2013. An updated Privacy Paradigm for the Internet of Things, available at: https://fpf.org/wp-content/uploads/Wolf-and-Polonetsky-An-Updated-Privacy-Paradigm-for-the-%E2%80%9CInternet-of-Things%E2%80%9D-11-19-2013.pdf [1 August 2017]

Raab, C.D., 2005. The future of privacy protection. Trust and Crime in Information Societies, p.14

Robinson, N., Graux, H., Botterman, M. and Valeri, L., 2009. Review of the European data protection directive. Cambridge: RAND, p.2

Schwartz, P.M. and Solove, D.J., 2011. The PII problem: Privacy and a new concept of personally identifiable information. NYUL rev., 86, p.1814

Schwartz, P.M. and Solove, D.J., 2013. Reconciling personal information in the United States and European Union, Cal. L. rev., 102, p.877 **["Schwartz & Solove, 2013, #1"]**

Schwartz, P.M. and Solove, D.J. 2013. Reconciling Personal Information in the United States and European Union, GW Law Faculty Publications & Other Works. Paper 956 **["Schwartz & Solove, 2013, #2"]**

Segrist, P., 2014. How the Rise of Big Data and Predictive Analytics Are Changing the Attorney's Duty of Competence. NCJL & Tech., 16, p.527

Simitis, S., 1999. Revisiting Sensitive Data. In Report of the Council of Europe, Review of the answers to the Questionnaire of the Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS 108), Strasbourg, p.1

Solove, D.J., 2002. Conceptualizing privacy. California Law Review, p.1087

Solove, D.J., 2007. I've got nothing to hide and other misunderstandings of privacy. San Diego L. Rev., 44, p.745

Sweeney, L., 2001. Computational disclosure control. A Primer on Data Privacy Protection

Tene, O., 2011. The complexities of defining personal data: anonymization. Data Protection Law Policy, 8, p.6

Tene, O. and Wolf, C., 2013. White Paper - The Definition of Personal Data: Seeing the Complete Spectrum, available at https://fpf.org/wp-content/uploads/FINAL-Future-of-Privacy-Forum-White-Paper-on-De-Id-January-201311.pdf [Accessed 1 August 2017]

Time.Lex, 2010, Study of case law on the circumstances in which IP addresses are considered personal data, SMART 2010/12 D3. Final report

UK Government, 2014. Emerging Technologies: Big Data - A Horizon Scanning Research Paper, available at:

www.gov.uk/government/uploads/system/uploads/attachment_data/file/389095/Horizon_Scanning_-_Emerging_Technologies_Big_Data_report_1.pdf [Accessed 1 August 2017]

Van Dijk, N., Gellert, R. and Rommetveit, K., 2016. A risk to a right? Beyond data protection risk assessments. Computer Law & Security Review, 32(2), p.286

Van den Hoven, J., 1999. Privacy and the varieties of informational wrong-doing. Australian Journal of Professional and Applied Ethics. 1, p.30

Van Den Hoven, J., Information technology, privacy and the protection of personal data. In Van Den Hoven, J. and Weckert, J. eds., 2008. Information technology and moral philosophy. Cambridge University Press, p.311

Varian, H.R., 2002. Economic aspects of personal privacy. In Cyber Policy and Economics in an Internet Age (pp. 127-137). Springer, Boston, MA

Westin, A.F., 1967. Privacy and freedom Atheneum. New York, 7

Wong, R., 2007. Data protection online: alternative approaches to sensitive data. J. Int'l Com. L. & Tech., 2, p.9

Wong, R., 2013. Data Protection in the Online Age, DOI: 10.2139/ssrn.2220754 [Accessed 1 August 2017]

Zwenne, G.J., 2013. Diluted Privacy Law – Paraphrased translation from Dutch of his inaugural lecture on 12 April 2013, available at: http://zwenneblog.weblog.leidenuniv.nl/files/2013/09/G-J.-Zwenne-Diluted-Privacy-Law-inaugural-lecture-Leiden-12-April-2013-ENG.pdf [Accessed 1 August 2017].

Reports

Banisar, D. and Davies, S., 2000. Privacy and human rights. Electronic Privacy Information Center (EPIC) (rev.2006), available at: https://sontusdatos.org/wp-content/uploads/2013/04/privacy_and_human_rights_2005-parte1.pdf [Accessed 1 August 2017]

Bellanova, R., Friedewald, M., Wright, D., Gellert, R., Gutwirth, S., Mordini, E., Schütz, P. and Vernier, S. 2011. Deliverable D.1 - PRESCIENT (Privacy and emerging fields of science and technology: Towards a common framework for privacy and ethical assessment), available at: http://www.prescient-project.eu/prescient/inhalte/download/PRESCIENT-D1---final.pdf [Accessed 1 August 2017]

Bibliography

Bennett, C.J., Haggerty, K.D., Lyon, D. and Steeves, V. eds., 2014. Transparent lives: surveillance in Canada. Athabasca University Press. [online] surveillanceincanada.org, available at: http://www.surveillanceincanada.org/ [Accessed 1 August 2017]

Cate, F.H. and Mayer-Schönberger, V., 2013 (rev.2014). Data Protection Principles for the 21st Century – Report. Oxford Internet Institute, available at: https://www.oii.ox.ac.uk/archive/downloads/publications/Data_Protection_Principles_for_the_21st_Century.pdf [Accessed 1 August 2017] **["Cate & Mayer-Schönberger, 2013, #1"]**

Centre for Information Policy Leadership (Hunter & Williams), 2014. The Role of Risk Management in Data Protection, Paper 2 of the Project on Privacy Risk Framework and Risk-based Approach to Privacy, available at: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_2-the_role_of_risk_management_in_data_protection-c.pdf/ [Accessed 1 August 2017]

Crosby, J., 2008. Challenges and opportunities in identity assurance. The Stationary Office Limited on behalf of HM Government, Treasury, London, UK

Dinant, J.M. and Poullet, Y, 2004. Report On The Application Of Data Protection Principles To The Worldwide Telecommunication Networks: Information Self-Determination In The Internet Era. In Thoughts On Convention No. 108 (Conseil de l'Europe, T-PD (2004) 04 final)

Dinant, J.M., 2010. Report on the lacunae of the Convention for the protection of individuals with regard to automatic processing of personal data (ETS No 108) resulting from technological developments (Part 1, Conseil de l'Europe, T-PD-BUR (2010) 09 (I) final)

European Commission. 2011, Privacy and Data Protection Impact Assessment Framework for RFID Applications, available at: http://cordis.europa.eu/fp7/ict/enet/documents/rfid-pia-framework-final.pdf [Accessed 1 August 2017]

European Data Protection Supervisor, 2015. Leading by Example – The EDPS Strategy 2015-2019, available at: https://edps.europa.eu/sites/edp/files/publication/15-07-30_strategy_2015_2019_update_en.pdf [Accessed 1 August 2017]

Friedewald, M., Wright, D., Gutwirth, S. and Mordini, E., 2015. Final Report Summary - PRESCIENT (Privacy and emerging fields of science and technology: Towards a common framework for privacy and ethical assessment), available at: http://cordis.europa.eu/result/rcn/155816_en.html [Accessed 1 August 2017]

Maude, F., 2012. Open Data White Paper - Unleashing the potential. The Stationary Office Limited on behalf of HM Government, Cabinet Office, London, UK

O'Hara, K., 2011. Transparent government, not transparent citizens: a report on privacy and transparency for the Cabinet Office. The Stationary Office Limited on behalf of HM Government, London, UK

House of Commons Science and Technology Committee. 2015. Current and future uses of biometric data and technologies. Sixth Report of Session 2014–15. The Stationary Office Limited on behalf of HM Government, London, UK, available at: https://publications.parliament.uk/pa/cm201415/cmselect/cmsctech/734/734.pdf [Accessed 1 August 2017]

House of Commons Science and Technology Committee. 2016. The Big Data Dilemma Fourth Report of Session 2015–16, HC 468. The Stationery Office on behalf of HM Government, London, UK, available at: https://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf [Accessed 1 August 2017]

UK Government Office for Science, 2013. Foresight Project Final Project report. Future Identities: Changing identities in the UK– the next 10 years, available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/273966/13-523-future-identities-changing-identities-report.pdf [Accessed 1 August 2017]

US Executive Office of the President and Podesta, J., 2014. Big data: Seizing opportunities, preserving values, available at: https://info.publicintelligence.net/WhiteHouse-BigDataReview.pdf [Accessed 1 August 2017]

Case-Law

<u>EU and other non-UK</u>

Asociacion Nacional de Establecimientos Financieros de Credito (ASNEF) and Federacion de Comercio Electronico y Marketing Directo (FECEMD) v. Administracion del Estado, Joined Cases C-468/10 and C-469/10, [2011] ECR I-12181

Belgian Commission for the Protection of Privacy v. Facebook Inc., Facebook Belgium SPRL and Facebook Ireland Limited 15/57/C

Bodil Lindqvist, Case C101-01, [2003] ECR I-12971

Bibliography

Breyer v. Bundesrepublik Deutschland, C-582/14, [2016] ECLI:EU:C:2016:779

College van burgemeester en wethouders van Rotterdam v. M.E.E. Rijkeboer, Case C-553/07, [2009] ECR I3889

Dyson v. Dormire, 2009 U.S. Dist. LEXIS 90211 (E.D. Mo. Sept. 1, 2009)

Digital Rights Ireland and Seitlinger and Others, Joined Cases C-293/12 and C-594/12, [2014] ECLI:EU:C:2014:238

EMI and Ors v. Eircom Ltd [2010] IEHC 108

Erich Stauder v. City of Ulm – Sozialamt, Case 29/69, [1969] ECR 419 303

GlaxoSmithKline Services and Others v. Commission, Joined Cases C-501/06P, C-515/06P and C-519/06P, [2009] ECR I-9291

Hoechst AG v. Commission, Joined Cases 46/87 and 227/88, [1989] ECR 2859

Nilsson and Others, Case C-162/97, [1998] ECR I-7477

Nowak v. Data Protection Commissioner, Case C-434/16, Opinion of Advocate General Kokott delivered on 20 July 2017, ECLI:EU:C:2017:582

Privacy Commissioner v. Telstra Corporation Limited [2017] FCAFC 4 CF

Rechnungshof v. Osterreichischer Rundfunk, Joined Cases C-465/00, C-138/01 and C-130/01, [2003] ECR I-04989

Regional Court Berlin, [2007] K&R, 532

Scarlet Extended SA v. Société belge des auteurs, compositeurs et éditeurs SCRL, Case C-70/10, [2011] ECR I-11959

Schrems v. Irish Data Protection Commissioner, Case C-362/14, [2015] 2015 ECR I-650

Sociéte d'Importatation Édouard Leclerc-Siplec v. TFI Publicité SA and M6 publicité SA, Case C-412/93, [1995] ECR 179

YS v. Minister voor Immigratie, Integratie en Asiel, Joined Cases C-141/12 and C-372/12, [2014] EU:C:2014:2081

UK

All Party Parliamentary Group on Extraordinary Rendition v. The Information Commissioner and the Ministry of Defence [2011] UKUT 153 (AAC)

Department of Health, R (on the application of) v. Information Commissioner [2011 EWHC 1430 (Admin) (20 April 2011)

Durant v. Financial Services Authority [2003] EWCA Civ 1746 (8 December 2003)

Efifiom Edem v. Information Commissioner and Financial Services Authority [2014] EWCA Civ 92 (7 February 2014)

Google Inc. v. Judith Vidal-Hall and others [2015] EWCA Civ 311 (27 March 2015)

Harcup v. Information Commissioner, Information Tribunal (5 February 2008)

Information Commissioner decision notice dated 23 January 2013, Reference: FS50514995

Information Commissioner decision notice dated 17 March 2014, Reference: FS50523095

Information Commissioner decision notice dated 27 October 2015, Reference: FS50565190

Ministry of Justice v. the Information Commissioner & Dr C Pounder EA/2012/0110

R (Kelway) v. The Upper Tribunal (Administrative Appeals Chamber) and Northumbria Police and R (Kelway) v. Independent Police Complaints Commission ([2013] EWHC 2575 (Admin), 20 August 2013)

Vidal-Hall and others v. Google Inc [2014] EWHC 13 (QB) (16 January 2014)

Technical Standards

Common Criteria for Information Security Evaluation. ISO/IS 15408: 1999

Health informatics – Pseudonymization. ISO/TS 25237:2008

Information Technology – Security Techniques. ISO/IEC 24760-1:2011

Books

Bamberger, K.A. and Mulligan, D.K., 2015. Privacy on the ground: driving corporate behavior in the United States and Europe. MIT Press.

Buchholz, R.A., 1992, 4th edition. Business environment and public policy: Implications for management. Englewood Cliffs, NJ: Prentice Hall

Collins Dictionary [online], available at: www.collinsdictionary.com/dictionary/english [Accessed 1 August 2017]

Floridi, L., Taylor, L, and Van der Sloot, B. eds., 2016. Group Privacy: New Challenges of Data Technologies. NYC: Springer

O'Hara, K. and Shadbolt, N, 2008. The Spy in the Coffee Machine. One World Publications

Oxford English Dictionary, 1989, 2nd edition, Oxford: Clarendon Press

Oxford English Dictionaries [online], available at: www.oxforddictionaries.com [Accessed 1 August 2017]

Hamilton, A. and Jay, R., 2012. 4th edition. Data protection law and practice, London: Sweet & Maxwell

2011. Consumers Anonymous?: The Privacy Risks of De-identified and Aggregated Consumer Data, Ottawa: Public Interest Advocacy Centre, available at: https://www.piac.ca/wp-content/uploads/2014/11/piac_consumers_anonymous_paper_final_6oct2011.pdf [Accessed 1 August 2017]

Meriam-Webster Dictionary [online], available at: http://www.merriam-webster.com/dictionary [Accessed 1 August 2017]

Phillips, J., Ryan, M.D., and Stalla-Bourdillon, S., 2014. Privacy vs. Security. NYC: Springer

Solove, D.J, 2008. Understanding Privacy. Cambridge, MA: Harvard University Press

Weber, R.H., 2003. Regulatory models for the online world, Leiden: Kluwer Law Intl

Letters, Press Releases and Speeches

Article 29 Working Party. Letter to Ms Ilze JUHANSONE Ambassador Extraordinary and Plenipotentiary Permanent Representative to the EU, Brussels, 17 June 2015, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150617_letter_from_the_art29_wp_on_trilogue_to_msjuhansone.pdf; and, ibid, Appendix, available at http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150617_appendix_core_issues_plenary_en.pdf [Both accessed 1 August 2017]

Article 29 Working Party. Letter to Mr Paul TIMMERS Director of Sustainable and Security Society Directorate DG Connect, Brussels, 25 February 2015, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_en.pdf; and, ibid, Appendix, available at http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf [Both accessed 1 August 2017]

Buttarelli, G. Antitrust, Privacy and Big Data. Brussels, 3 February 2015, [online], available at: https://edps.europa.eu/sites/edp/files/publication/15-02-03_competition_big_data_speech_gb_en.pdf [Accessed 1 August 2017]

Denham, E. GDPR and Accountability. London, 17 January 2017, [online], available at: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/01/gdpr-and-accountability/ [Accessed 1 August 2017]

European Commission (2010). Data protection: Commission requests UK to strengthen powers of national data protection authority, as required by EU law. [online], 24 June 2010, available at: http://europa.eu/rapid/press-release_IP-10-811_en.htm [Accessed 1 August 2017]

European Commission (2017). Questions and Answers – Data Protection Reform Package [online], 24 May 2017, available at: http://europa.eu/rapid/press-release_MEMO-17-1441_en.htm [Accessed 1 August 2017]

Hustinx, P. Protection of personal data on-line: the issue of IP addresses, 15 April 2009, [online], available at: https://secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/EDPS/Publications/Speeches/2009/09-04-15_adresses_IP_EN.pdf [Accessed 1 August 2017]

Information Commissioner's Office. (2012). New anonymisation code sets out how to manage privacy risks and maintain transparency. [online], 20 November 2012, available at: http://www.ico.gov.uk/news/latest_news/2012/new-anonymisation-code-sets-out-how-to-manage-privacy-risks-and-maintain-transparency-20112012.aspx [Accessed 1 August 2017].

Reding, V., The EU data protection Regulation: Promoting technological innovation and safeguarding citizens' rights, 4 March 2014, [online], available at: http://europa.eu/rapid/press-release_SPEECH-14-175_en.htm [Accessed 1 August 2017]

Online Articles

Baines, J. (2015). The Wrong Test for Anonymisation. [online] information rights and wrongs, available at: http://informationrightsandwrongs.com/2014/12/09/the-wrong-test-for-anonymisation/ [Accessed 1 August 2017]

Cheshire, T. (2017). Tfl plans to make £322m by collecting data from passengers' mobiles via Tube Wi-Fi. [online] Sky News, available at: http://news.sky.com/story/tfl-may-make-322m-by-selling-on-data-from-passengers-mobiles-via-tube-wifi-11056118 [Accessed 1 October 2017]

Cronk, R. (2016). How your legal background may work against you. [online] IAPP, available at: https://iapp.org/news/a/how-your-legal-background-may-work-against-you/ [Accessed 1 August 2017]

Davis, W. (2015). Supercookies, Digital Fingerprinting Undermine Trust In Web, W3C Says. [online] media.post.com, available at: https://www.mediapost.com/publications/article/254275/supercookies-digital-fingerprinting-undermine-tru.html?print [Accessed 1 August 2017]

Finch, K. (2014). The Evolving Nature of Consumer Privacy Harm [online] privacy association, available at: https://www.privacyassociation.org/publications/the_evolving_nature_of_consumer_privacy_harm [Accessed 1 August 2017]

Johnston, A. (2017). Mobiles, metadata and the meaning of 'personal information. [online] salinger privacy, available at www.salingerprivacy.com.au/2017/01/19/federalcourtdecision/ [Accessed 1 August 2017]

Knight, A. (2015). Court of Appeal upholds landmark judgement against Google arising from its exploitation of Apple's Safari web-browser privacy settings. [online] peepbeep, available at: https://peepbeep.wordpress.com/2015/04/04/court-of-appeal-upholds-landmark-judgement-against-google-arising-from-its-exploitation-of-apples-safari-web-browser-privacy-settings/ [Accessed 1 August 2017]

Knight, A. (2016). Latest Policy Guidance Published on Data Protection and Location Analytics Data. [online] peepbeep, available at: https://peepbeep.wordpress.com/2016/03/11/latest-policy-guidance-published-on-data-protection-and-location-analytics-data/ [Accessed 1 August 2017]

Knight, A. (2016). Mind the Caveats – CJEU Advocate General opines that Dynamic IP Addresses can be Personal Data … (sometimes). [online] peepbeep, available at: https://peepbeep.wordpress.com/2016/06/17/mind-the-caveats-cjeu-advocate-general-opines-that-dynamic-ip-addresses-can-be-personal-data-sometimes/ [Accessed 1 August 2017]

Knight, A. (2017). Advocate General Delivers Opinion on Whether Examination Scripts Are Personal Data under Data Protection Law. [online] peepbeep, available at: https://peepbeep.wordpress.com/2017/07/27/2885/ [1 August 2017]

Knight, A. (2017). ICO requests feedback on new data protection profiling provisions [online] peepbeep, available at: https://peepbeep.wordpress.com/2017/04/10/ico-requests-feedback-on-new-data-protection-profiling-provisions/ [1 August 2017]

Knight, A. (2017). New EU guidelines on data protection impact assessments [online] peepbeep, available at: https://peepbeep.wordpress.com/2017/04/18/new-eu-guidelines-on-data-protection-impact-assessments/ [Accessed 1 August 2017]

Knight, A. & Stalla-Bourdillon, S. (2016). The First-Tier Tribunal and the anonymisation of clinical trial data: a reasoned expression of Englishness…. which would have to be abandoned with the GDPR? [online] peepbeep, available at: https://peepbeep.wordpress.com/2016/09/19/the-first-tier-tribunal-and-the-anonymisation-of-clinical-trial-data-a-reasoned-expression-of-englishness-which-would-have-to-be-abandoned-with-the-gdpr/ [Accessed 1 August 2017]

Lei, T. (2015). Singapore and UK researchers investigate private in big data era. [online] computer weekly, available at: http://www.computerweekly.com/news/4500254910/Singapore-and-UK-researchers-investigate-privacy-in-big-data-era [Accessed 1 August 2017]

Pinsent Masons (2008). IP addresses and the Data Protection Act? [online] out-law, available at: www.out-law.com/en/topics/tmt--sourcing/data-protection-and-privacy/ip-addresses-and-the-data-protection-act/ [Accessed 1 August 2017].

Stalla-Bourdillon, S. (2015). The Article 29 Working Party on the concept of health data: could it mean that we need to adapt the definition of health data as well as that of personal data? [online] peepbeep, available at: https://peepbeep.wordpress.com/2015/02/10/article-29-working-party-on-the-concept-of-health-data-could-it-mean-that-we-need-to-adapt-the-definition-of-health-data-as-well-as-that-of-personal-data/ [Accessed 1 August 2017]

Stalla-Bourdillon, S. (2016). A call for a common techno-legal language to speak about anonymisation, pseudonymisation, de-identification… Could this be one of the biggest challenges brought about by the GDPR? [online]) peepbeep, available at:

https://peepbeep.wordpress.com/2016/11/09/a-call-for-a-common-techno-legal-language-to-speak-about-anonymisation-pseudonymisation-de-identification-could-this-be-one-of-the-biggest-challenges-brought-about-by-the-gdpr/ [Accessed 1 August 2017]

Tene, O. (2015). Privacy Is the New Antitrust: Launching the FTC Casebook. [online] IAPP, available at: https://iapp.org/news/a/privacy-is-the-new-antitrust-introducing-the-ftc-casebook/ [Accessed 1 August 2017]

Case study background information (Chapters 1 and 4 references)

Denham, E. Letter to Sir David Solma, 3 July 2017, available at https://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf [Accessed 12 August 2017]

Hearn, A. Google DeepMind pairs with NHS to use machine learning to fight blindness. [online] The Guardian, 5 July 2016, available at: https://www.theguardian.com/technology/2016/jul/05/google-deepmind-nhs-machine-learning-blindness [Accessed 1 August 2017]

Hearn, A. Google DeepMind and UCLH collaborate on AI-based radiotherapy treatment. [online] The Guardian, 30 August 2016, available at: https://www.theguardian.com/technology/2016/aug/30/google-deepmind-ucl-ai-radiotherapy-treatment- [Accessed 1 August 2017]

Information Commissioner's Office. (2017). Royal Free - Google DeepMind trial failed to comply with data protection law. [online], 3 July 2017, available at: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/ [Accessed 1 August 2017]

King, D. & Suleyman, M. (2017). The Information Commissioner, the Royal Free, and what we've learned. [online] DeepMind, 3 July 2017, available at: https://deepmind.com/blog/ico-royal-free [Accessed 1 August 2017]

Powles, J. and Hodson, H., 2017. Google DeepMind and healthcare in an age of algorithms. Health and technology, 7(4), p.351.

Royal Statistical Society, 2014. Royal Statistical Society research on trust in data and attitudes toward data use / data sharing, available at: https://www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf [Accessed 1 August 2017]