Can a privacy-based approach help regulate secondary uses of patients' data?

Miranda Mourby

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Humanities
School of Law

July 2025

Acknowledgements

First and foremost, I would like to thank my PhD supervisors Dr Jiahong Chen, Professor Sara Fovargue and Professor Naomi Hawkins, for their careful and patient review of my work over the last three years.

My doctoral research was made possible by a PhD scholarship from the School of Law at the University of Sheffield, for which I am very grateful.

Additionally, as this thesis is in Publication Format, it contains seven publications which were supported (in whole or in part) by other project funding. The relevant funding is acknowledged at the start of each publication, but collectively I would like to acknowledge:

- The Economic and Social Research Council grant ES/L007452/1 for the Administrative Data Service;
- The Leverhulme Trust project grant RPG-2017-330 'Biomodifying Technologies: Governing converging research in the life sciences;
- The Medical Research Council project grant MC_PC_17215 'Developing an informatics supported platform for experimental medicine - MICA, industrial collaboration with Janssen and SomaLogic';
- The EU-STANDS4PM consortium (www.eustands4pm.eu) that was funded by the European Union Horizon2020 framework programme of the European Commission under Grant Agreement #825843;
- the Bundesministerium für Bildung und Forschung project awarded to Professor Fruzsina Molnár-Gábor: Datenschutzrechtliches Reallabor für eine Datentreuhand in der Netzwerkmedizin;
- The Medical Research Council project grant for the UK BioLink project;
- The Innovative Medicines Initiative Joint Undertaking under grant agreement no115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution.

I am very grateful to all the co-authors with whom I collaborated in the preparation of the articles and book chapters contained in this thesis. They are named in the title page of each publication, but altogether they are (in publication order): Dr Elaine Mackey, Professor Mark Elliot, Dr Heather Gowans, Dr Susan E Wallace, Dr Jessica Bell, Dr Hannah Smith, Professor Stergios Aidinlis, Professor Jane Kaye and Professor Fruzsina Molnár-Gábor.

I am also indebted to the reviewers and editors who provided feedback on my draft publications. In particular: Professor Steve Saxby, Professor Mark Taylor, Dr Bart van der Sloot, Dr Sascha van Schendel and Professor Edward S Dove.

On a personal note, I am grateful to Dr Katharine Nicholas and Professor Ritchie Robertson for their support, in this project and many others. I would like to dedicate this completed work to Nik and Tom.

Abstract

In this thesis, I establish whether, and how, a ‘privacy-based’ approach can help regulate secondary uses of patients’ data. I use Article 8 of the European Convention on Human Rights as the bedrock of my definition of ‘privacy.’ My approach thus re-grounds the law governing patients’ data in a foundational legal standard, which cuts across multiple strands of legal doctrine.

I address this question over the course of seven publications, through which run the threads of five case studies of secondary use: scientific research, tissue donation, immigration, software development and public health analytics. The result of my analysis is a clarified conception of three key aspects of Article 8, which help determine the scope and nature of patients’ rights in their data:

- Identification;
- Private Life;
- Justification.

This thesis is structured in three Parts, with each Part centring on one of these concepts in turn. Parts 1 and 2 respectively define ‘identification’ and ‘private life’ as relating to information which can interfere with Article 8. Part 3 then considers how secondary uses of patients’ data should be justified under Article 8, with particular emphasis on the requirements of proportionality and non-discrimination.

Ultimately, I argue that these clarified conceptions of identification, private life and justification can help govern secondary uses of patients’ data. They help interpret data protection, confidentiality and privacy law in a way that not only creates greater intellectual coherence, but also improves legal protection for patients. I conclude that all secondary uses of identifiable patient data engage Article 8, and thus require robust, systematic justification.

List of Contents

Declaration

I, the author, confirm that the Thesis is my own work. Where any publications contained within this thesis have been co-authored, the contribution of my co-authors is explained in the Synopsis to the relevant Part of the thesis.

I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously presented for an award at this, or any other, university.

Abbreviations

A29WP	Article 29 Working Party
DPIA	Data Protection Impact Assessment
DPIA+	A Data Protection Impact Assessment, supplemented by consideration of the Public Sector Equality Duty (my coinage)
ECHR	European Convention on Human Rights
ECJ	European Court of Justice
ECtHR	European Court of Human Rights
EDPB	European Data Protection Board
EHDS	European Health Data Space
GDPR	General Data Protection Regulation
HSCA	Health and Social Care Act 2012
MOPI	Misuse of Private Information
MOU	Memorandum of Understanding
NHS	National Health Service
NHSA	National Health Service Act 2006
PSED	Public Sector Equality Duty
SoS	Secretary of State

Table of Cases

ENGLAND AND WALES

R (Mahmood) v Secretary of State for the Home Department [2001] 1 WLR 840

R (MXK, EH, HH, SXB, and ALK) v Secretary of State for the Home Department [2023] EWHC 1272 (admin)

R (Public Law Project) v Lord Chancellor [2016] UKSC 39

R (on the application of Spath Holme Ltd) v Secretary of State for the Environment, Transport and the Regions [2001] 2 AC 349

R (on the application of W, X, Y, and Z) v The Secretary of State for Health v The Secretary of State for the Home Department [2014] EWHC 1532 (admin)

R (on the application of W, X, Y, and Z) v The Secretary State for Health v The Secretary of State for the Home Department, The British Medical Association [2015] EWCA Civ 1034

R (Source Informatics Ltd) v Department of Health [1999] EWCA Civ 3011

South Lanarkshire Council v The Scottish Information Commissioner [2013] UKSC 55

W v Egdell [1989] EWCA Civ 13

Weller & Ors v Associated Newspapers Ltd [2015] EWCA Civ 1176

EUROPEAN COURT OF HUMAN RIGHTS

Airey v Ireland (1979-80) 2 EHRR 305

Antović and Mirković v Montenegro [2017] 11 WLUK 675

Bărbulescu v Romania| [2017] 9 WLUK 42

Copland v The United Kingdom [2007] ECHR 253

Glass v The United Kingdom (2004) 39 EHRR 15

Halford v The United Kingdom (1997) EHRR 523

*Handyside v The United Kingdom (*1979-80) 1 EHRR 737

Julien v France (1991) App. No(s).14461/88, 9 July 1991

Karin Köpke v Germany (2011) 53 EHRR. SE26

Magyar Helsinki Bizottság v Hungary (2020) 71 EHHR 2

MM v The United Kingdom [2012] 11 WLUK 360

Mustafa Erdogan and Ors v Turkey App Nos. 346/04 and 39779/04, 27 May 2014

Pay v The United Kingdom (2009) 48 EHRR SE2

Peck v The United Kingdom (2003) 36 EHRR 41

Perry v The United Kingdom (2008) 41 EHRR 1007

PG & JH v The United Kingdom (2008) 46 EHRR 51

S and Marper v The United Kingdom (2009) 48 EHRR 50

Sunday Times v The United Kingdom (1979-80) 2 EHRR 245

Vicent del Campo v Spain [2018] ECHR 909

Von Hannover v Germany (2005) 40 EHHR 1

Z v Finland (1998) 25 EHRR 371

COURT OF JUSTICE OF THE EUROPEAN UNION

Case C 274/99 *Connolly v Commission* [2001] ECR I-1611

Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* | [2017] 2 CMLR 3

Case C-465/00 *Rechnungshof v Österreichischer Rundfunk* [2003] ECR I-4989

Case T- 557/20 *Single Resolution Board v European Data Protection Supervisor* [2024] 1 CMLR 46

Table of Statutes

ENGLAND AND WALES

Control of Patient Information Regulations 2002

Data Protection Act 2018

Digital Economy Act 2017

Health and Social Care Act 2012

National Health Service Act 2006

EUROPEAN UNION

Charter of Fundamental Rights of the European Union

Council Directive 2001/83/EC of 6 November 2001 on the Community code relating to medicinal products for human use

Council Directive 2004/23/EC of 31 March 2004 on setting standards of quality and safety for the donation, procurement, testing, processing, preservation, storage and distribution of human tissues and cells

Council Directive (EU) 2010/45 of 7 July 2010 on standards of quality and safety of human organs intended for transplantation

Council Regulation (EC) 1394/2007 of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004

Council Regulation (EU) 536/2014 of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC

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

Explanatory Memorandum, Draft Charter of Fundamental Rights of the European Union, Convention document CHARTE 4473/00, 11 October 2000

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

Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847

COUNCIL OF EUROPE

Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 1950 Council of Europe European Treaty Series 5

Council of Europe Convention of 1981 for the protection of individuals with regard to automatic processing of personal data.

COMMITTEE REPORTS

Committee on Civil Liberties, Justice and Home Affairs of the European Parliament, *Working Document 2 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)* (2009 – 2014; 8 October 2012)

Council of Europe's T-PD Committee, *Report on the application of Data Protection Principles of the Worldwide Telecommunication Network,* (2004) 04 final)

Health and Social Care Committee, *Memorandum of understanding on data-sharing between NHS Digital and the Home Office* (2017-9, HC 677)

Joint Committee on Human Rights, *Legislative Scrutiny: Offender Management Bill (third report)* (2006-7), Appendix 6) (Q3)

EUROPEAN DATA PROTECTION BOARD

Guidelines 01/2025 on Pseudonymisation (Brussels, 16 January 2025)

Guidelines WP 260 on transparency under Regulation 2016/679 (Brussels, 11 April 2018)

Opinion 05/2014 on Anonymisation Techniques (Brussels, 10 April 2014)

Opinion 4/2007 on the concept of personal data (Brussels, 20 June 2007)

Introduction

I Research Question

The research question of this thesis asks whether a 'privacy-based approach' can help regulate secondary uses of patients' data in England. I will begin by explaining why this question is worth addressing. That is, why are there issues within secondary uses of patients' data that require the 'help' of a particular approach? The issues I have identified are partly legal doctrinal in nature— a disjointed heterogeneity in the relevant legal framework that leads to a lack of clarity. But there is also a deeper, more normative issue which I will explore: the need for a central, underpinning principle to protect patients' interests in their information. This thesis therefore asks whether a robust and consistent use of the right to private life under Article 8 of the European Convention on Human Rights ('ECHR')[1] as an interpretive benchmark can not only create better doctrinal coherence, but also better protection for patients as individuals and groups across society.

The next subsections of this introduction will therefore discuss the policy background to my research question, as well as the issues within the legal status quo. I will then explain, in more detail, how my 'privacy-based approach' is based in the text of Article 8 ECHR and its associated jurisprudence, and how this can help ground evaluation of secondary uses of patients' data in the potential for interference with their private life (i.e. dignity,[2] autonomy[3] and close relationships[4]).

The aim of my re-evaluation is a clarified conception of three main aspects of Article 8 ECHR:

1) Identification
2) Private Life
3) Justification

These concepts underpin multiple areas of English information law, most significantly: data protection, confidentiality and misuse of private information ('MOPI'). It is the ultimate argument of this thesis that a clarified understanding of these concepts can help regulate secondary uses of patients' data. This 'help' will be understood in terms of promoting

---

[1] Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 1950 Council of Europe European Treaty Series 5.
[2] Dignity is a core part of Article 8 ECHR, but arguably also a foundational principle of the ECHR generally. See C. Heri, 'Deference, Dignity and 'Theoretical Crisis': Justifying ECtHR Rights Between Prudence and Protection' (2024) 24 Human Rights Law Review 1, 1-19.
[3] See 'Beyond Autonomy,' Section VI C below.
[4] See, for example, *S and Marper v The United Kingdom* (2009) 48 EHRR 50, [66]: 'The Court notes that the concept of "private life" is a broad term not susceptible to exhaustive definition. It covers the physical and psychological integrity of a person […]. Beyond a person's name, his or her private and family life may include other means of personal identification and of linking to a family.'

doctrinal coherence, as well as the policy implications for secondary uses of patients' information.

As I am focusing on secondary uses of patient data in England,[5] the next subsection explains the significance of these uses within the English National Health Service ('NHS'), and why there is a policy need to re-evaluate the relevant law using my 'privacy-based' approach.

II Policy Context

A. What are 'secondary' uses of patients' data?

The idea of a 'secondary' use of data spans data science, public policy and (more recently) law. In its broadest sense, 'secondary' use simply means a use of information other than the 'primary' one for which it was originally collected. My focus in this thesis is on the re-use of NHS patient data collected initially for the primary purpose of healthcare delivery in England.

Background

The term 'secondary use' of data gained initial significance in the dawn of computing, when the collection, aggregation and re-use of digitised information began to yield new insights for governments sitting on large amounts of data. Daniel Solove traces the modern use of the term back to a 1973 report from the US Department of Health, Education, and Welfare on the harms caused by computer databases.[6] This report led to a 'purpose specification principle' within US law, limiting the range of unconsented secondary uses for which agencies could process the data they had acquired.[7]

Secondary uses of data have gained renewed potency in the 'Big Data' era, in which further advances in computing have expanded the volume and variety of data that can be analysed through automation.[8] The dynamics of 'Big Data,' in which information is amassed for a potential multiplicity of future uses, come into an inevitable tension with data protection's purpose limitation principle,[9] even though some secondary uses (such as scientific research) benefit from a presumption of compatibility under the General Data Protection Regulation.[10]

---

[5] The other NHS systems within the UK have their own governing legislation, and have not reported the same backlash to their secondary uses of health data. For legal and policy reasons, therefore, I am focusing on England in this thesis.
[6] D. J Solove, 'A taxonomy of privacy' (2006) 154 University of Pennsylvania Law Review 3, 477–564, 518.
[7] Ibid, 519.
[8] D. R. Schlegel and G. Ficheur, 'Secondary Use of Patient Data: Review of the Literature Published in 2016' (2017) 26 Yearbook of Medical Informatics 1, 68-71.
[9] N. Forgó, S. Hänold and B. Schütze, 'The Principle of Purpose Limitation and Big Data Perspectives in Law, Business and Innovation' in M. Corrales, M. Fenwick and N. Forgó (eds) *New Technology, Big Data and the Law* (2017) (Singapore: Springer), 17-42.
[10] 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 (hereafter 'the GDPR'), Article 5.1(b). The scope of the scientific research exemptions is considered in more

Secondary Uses of Patient Data

In this thesis, I am particularly concerned with the secondary use of information collected by the NHS, which is re-use for purposes other than delivering healthcare. The secondary uses considered within this thesis include research, tissue donation, immigration monitoring, algorithmic development and public health analytics. As disparate as they may seem, their commonality lies in the fact that a patient had information collected for a healthcare purpose, and then the information was re-used in another way.

The importance of secondary uses of health data is now so widely recognised that the EU has introduced a new legal architecture for cross-border processing of health data for these multiple re-uses. The Proposal for the Regulation on the European Health Data Space ('EHDS') defines secondary uses of electronic health data with reference to eight categories of use, including scientific research, medical training, testing algorithms and to assist public bodies in carrying out their functions.[11] The way I use the term 'secondary uses' of patient data in this thesis is relatively simple, and spans many of the above categories. My own working definition of 'secondary use,' here, is *any use of NHS patient data, other than for the delivery of care to that individual.*

I am deliberately using the term 'patient data,' not 'health data.' This is to make it clear that I am interested in *all* the information collected by the NHS from its patients which is then re-used for another purpose. As I will explore in Part 2 of this thesis, the current statutory focus on the confidentiality of health-related information—and the corresponding suggestion that merely 'demographic' information is not confidential[12]— does not go far enough, and creates holes in the legal protection afforded to patients. These loopholes can seem arbitrary— for example, if a patient's name and address are not taken from their medical records, it may not be considered 'confidential patient information,'[13] despite the majority of English patients who do believe their address should be confidential.[14] In my view, the administrative origin of the information should not be the deciding factor, but rather its potential to interfere with a patient's private life.

---

detail in the 5.academic governance article **(thesis page 75).** Unless specified otherwise, references to 'the GDPR' in this thesis are to both the UK and EU GDPR, as the relevant text is substantively the same.

[11] Regulation 2025/327 on the European Health Data Space, hereafter cited as 'the EHDS Regulation,' Article 2(e) and Article 34.

[12] See for example, NHS Digital, 'Patient data and confidential patient information' (12 October 2023) https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out/confidential-patient-information.

[13] Under s.251(10) NHS Act 2006, 'patient information' must either relate to health, or must be derived directly or indirectly from information relating to health, care and treatment. See H. Evans, 'Using data in the NHS: the implications of the opt-out and GDPR' (24 May 2018) https://www.kingsfund.org.uk/insight-and-analysis/long-reads/using-data-nhs-gdpr.

[14] Academy of Medical Royal Colleges, 'Disclosing personal demographic data: The public interest. Report of a seminar held at the Academy of Medical Royal Colleges' (June 2021) https://www.aomrc.org.uk/wp-content/uploads/2021/06/Disclosing_personal_demographic_data_0621.pdf, 9.

The central concern of this thesis is, therefore, any information recorded about an NHS patient for individual care,[15] which is then re-used for another purpose. There will, naturally, also be legal issues surrounding the secondary use of other kinds of health data— for example, re-using data from clinical trials, or from commercial mobile health apps. But secondary uses of patient data from England's NHS are unique in terms of scale, national significance, and controversy. I shall explain this further in the next subsection.

B. Why examine secondary uses of patient data?

Secondary uses of patient data are vitally important for the NHS. They are needed to develop new treatments, evaluate existing services and maintain its financial functioning through invoice validation,[16] to name but a few reasons. But the legitimacy of these large-scale re-uses should not be taken for granted. Put simply, the NHS in England has both a significant interest in pursuing secondary uses of health data, and a history of failures in this pursuit.

Care.data

The most infamous failure, at least within academic literature,[17] is NHS England's 'care.data' programme, which was abandoned in 2016 following backlash from doctors and the wider public. Even before this, however, a pattern of failed attempts to centralise NHS patient data was established. The 2002-2011 NHS Programme for IT also went over budget, and ran into objections from patients and clinicians due to its top-down implementation.[18] The 'Big Data' model of secondary uses inevitably creates a degree of tension: patients' data are a more potent tool for the government when amassed and analysed at a national level, but this nationalisation can come at the cost of the local autonomy of patients and their healthcare providers.

The lessons from care.data have not prevented some repetition of the patterns that drove its failure: top-down mandates, short timeframes for implementation, and a lack of clarity for patients about how their information would be used. In 2021, the planned GP Data for Planning and Research ('GPDPR') database was paused following concerns that the parameters of the data-sharing, and the corresponding opt-out, were insufficiently clear to the

---

[15] Which could include e.g. the demographic information which must be provided to register with a healthcare provider.

[16] Although invoice validation is essential for the continued flow of funds from NHS commissioners to providers, it is authorised ad-hoc through extensions under s.251 NHS Act 2006. See NHS England, 'Invoice validation' https://www.england.nhs.uk/ig/in-val/.

[17] R. Hays and G. Daker-White, 'The care.data consensus? A qualitative analysis of opinions expressed on Twitter' (2015) 15 BMC Public Health 1, 838; S. Sterckx., V. Rakic, J. Cockbain and P. Borry, '"You hoped we would sleep walk into accepting the collection of our data": controversies surrounding the UK care.data scheme and their wider relevance for biomedical research' (2016) 19 Medicine, Health Care and Philosophy 2, 177-190; L. Presser, M. Hruskova, H. Rowbottom and J. Kancir, 'Care.data and access to UK health records: patient privacy and public trust' (2015) Technology Science, 2015081103.

[18] T. Justinia, 'The UK's National Programme for IT: Why was it dismantled?' (2017) 30 Health Services Management Research 1, 2-9.

patient population.[19] A mere two years later, in 2023, the NHS Federated Data Platform—a next generation of centralised patient data—was challenged in a letter before action, on the grounds that its proposed data flows were unlawful.[20]

The 'Social Licence' for Secondary Uses

The consistent run of setbacks and controversies outlined above raises a policy question as to how secondary uses of patient data can be legitimate, acceptable and trusted. Some have understandably concluded that the fundamental social licence for secondary uses of health data is unclear,[21] meaning the issue lies in a deeper lack of public consensus, rather than any legal doctrine. This lack of consensus appears to be a distinctly English problem, with the other UK devolved nations not reporting comparable backlashes from their patient populations.[22] Understandably, therefore, some academic lawyers have called for remedies from outside the law to resolve the controversy around secondary uses of data. These have included public deliberation,[23] or participatory models of governance that re-centre patients' voices.[24] These participatory solutions have had some success in resolving the backlash to the care.data[25] and the GPDPR database.[26] That said, none of these participatory efforts have thus far prevented the recurring backlash and legal ambiguity between 2002 and 2023 (see the care.data narrative on the previous page). There is a noted risk that attempts to involve the public may become tokenistic acts of 'window dressing'[27] within a public-sector bureaucracy.

Ultimately, this is not a thesis about the merits or limitations of public engagement. It is about whether a privacy-based approach can clarify and strengthen the legal protection of patients' interests when their data are re-used for secondary purposes, rather than how public support for these uses can be cultivated. In reflecting on when secondary uses of patients' data can be justified as 'necessary in a democratic society,'[28] I will use Article 8 as a benchmark, rather

---

[19] G. Melvin, 'We must listen to the public on GP data' (26 May 2022) https://digital.nhs.uk/blog/data-points-blog/2022/we-must-listen-to-the-public-on-gp-data.

[20] Foxglove, 'Legal action launched: no legal basis for the £330 million Palantir NHS Federated Data Platform' (30 November 2023) https://www.foxglove.org.uk/2023/11/30/legal-action-palantir-nhs-federated-data-platform/. From my own enquiries, it seems this initial threat of litigation did not go any further, but has been kept under review.

[21] P. Carter, G. T. Laurie and M. Dixon-Woods, 'The social licence for research: why care.data ran into trouble' (2015) 41 J Med Ethics 5, 404-9.

[22] M. McCartney, 'Care.data: why are Scotland and Wales doing it differently?' (2014) 20 British Medical Journal 348, 1702; B. O'Brien, 'Care.data: how Northern Ireland is doing it' (2014) 3 British Medical Journal 348, 2380.

[23] E.g. 'collaborative public reasoning' per M. J. Taylor and J, Wilson, 'Reasonable Expectations of Privacy and Disclosure of Health Data' (2019) 27 Medical Law Review 3, 432–460, 460.

[24] J. Kaye, S. F. Terry, E. Juengst et al, 'Including all voices in international data-sharing governance' (2018) 12 Human Genomics 13. However, Iusmen and Boswell (note 27) express scepticism over the incorporation of this model into NHS England's governance.

[25] Sterckx et al, note 17.

[26] Carter et al, note 21.

[27] I. Iusmen and J. Boswell, 'The Dilemmas of Pursuing "Throughput Legitimacy" through Participatory Mechanisms' (2016) 40 West European Politics 2, 459–78.

[28] Per ECHR, Article 8(2).

than any empirical data of public attitudes (see the explanation of my methodology at section V of this Introduction).

Research Question on Secondary Uses

The 'launch, pause, consult' pattern which has for years dogged secondary uses of patient data in England suggests a disconnect between (the operation of) State powers to re-use patients' data, and the wider social and political expectations of the patients who are the subjects of this information. Against the backdrop of this tension, I will re-evaluate the laws governing secondary uses of patient data in England, and consider whether they sufficiently reflect the interests contained in Article 8 ECHR. This provision represents a fundamental, constitutional standard regulating State interference in private lives, against which a national legal framework can be judged. In this regard, this thesis aims to address the following overarching research question:

Can a privacy-based approach help regulate secondary uses of patients' data?

This question will be answered through careful exploration of five case studies, across seven publications. The case studies I will examine indicate that NHS information governance can become opaque, rigid and bureaucratic, and— ultimately— closed off to considerations of how different patients are impacted when their data are re-used. A legal doctrinal response to this issue is to re-centre Article 8 ECHR in the interpretation of the relevant legal framework. Article 8 serves as my key benchmark both for defining the *scope* of patients' rights within this framework, and for how interference with these rights should be *justified*. As well as bringing coherence to a fragmented statutory/common law system, this privacy-based approach can make space for consideration of the patient in their full, non-quantified humanity, by prompting questions of dignity, autonomy, and non-discrimination.

The next section provides an overview of the legal architecture underpinning NHS England's secondary uses of patient data, providing more context for why my re-evaluation via a 'privacy-based approach' is warranted. It sets out this legal framework in the context of the case studies explored across the publications in this thesis.

III. Case Studies & Legal Framework

A. How does English law regulate secondary uses of patient data?

The previous section provided, by way of background, a history of the care.data programme, and subsequent initiatives for secondary data use that have either failed, or at least been met with significant backlash. This section now focuses on the case studies which feature in the publications included in this thesis. The case studies are outlined in the context of the current legal framework for secondary uses of patients' data. As this legal framework is not set out in any single publication in this thesis, it is summarised here by way of overview.

Case Studies of Secondary Use

In preparing this thesis, I have identified five examples of the types of secondary uses of NHS patient data. These are discussed across the seven publications included in this thesis. To highlight their recurrence across the thesis, the integrative text refers to these **Case Studies** using **Bold Underlining**. The aim of this formatting choice is to make it visually clearer that these types of secondary use do, in fact, recur throughout the thesis, as key illustrations of the harms and benefits of re-using patients' data.

The thesis publications are referred to by number and (shortened) name, and are underlined. This formatting in the integrative text is intended to highlight the signpost to different publications in this thesis. Being a thesis in publication format, some additional signposting has been included, to attempt greater coherence across its integrated parts.

By way of overview, the table on the next page offers a visual 'map' of the publications and case studies of this thesis:

Publication Number	Publication Short Title	**Case Study**	Key Aspect of Article 8
1	1. Pseudonymisation article	This journal article focuses on whether pseudonymised data can be anonymised for research. Anonymous use of data is a key aspect of the **Scientific Research Study.**	Identification
2	2. Anonymity article	This journal article makes a case for preserving the category of legally anonymous data, both within **Scientific Research** and the **Tissue Donation Case Study**.	Identification
3	3. Profiling chapter	This book chapter considers, in more depth, the concept of 'identity' within identification, with further reference to **Scientific Research**.	Identification
4	4. Reasonable Expectations article	This journal article argues that secondary uses of patients' data should be considered private, by default, with reference to the **Immigration Case Study** and the **DeepMind Case Study**.	Private Life
5	5. Academic governance article	This journal article considers the impact of Article 8 on the **Scientific Research** exemptions of the GDPR, and how privacy rights can help close up any 'loopholes' in patient protection.	Justification
6	6. EHDS chapter	This book chapter takes an EU-wide perspective on the exemptions for **Scientific Research**, and considers how much control patients should have over their data under Article 8 ECHR.	Justification
7	7. DPIA+ article	This journal article explores the additional evaluative requirements of Article 8 for the GDPR's Data Protection Impact Assessment, in the context of the **Covid-19 Case Study.**	Justification

Table 1: Thesis Map

As the above table indicates, the **Scientific Research Case Study** recurs most throughout this thesis. It features as the predominant example of the benefits of secondary uses of patients' data, and my publications explore both its intended aims, and its potential for excessive intrusion into patients' privacy. This case study is a particular focus in the context of 'Identification' (Part 1 of this thesis), given the importance of anonymous data for scientific research. In Part 3 of this thesis ('Justification'), the focus is more how Article 8 ECHR can limit any disproportionate reliance on the GDPR's scientific research exemptions.

The **Tissue Donation Case Study** is an additional example of the benefits of secondary use of patients' data. In this scenario, donors' data *are* used for healthcare purposes, but for the care of a different patient, and thus still fall within my broad definition of 'secondary uses.' This case study is reviewed within the 'Identification' focus of Part 1, as illustrating the case for preserving donor anonymity to an appropriate legal standard.

The other case studies— **DeepMind**, **Immigration** and **Covid-19**— feature more in Parts 2 and 3 of this thesis, and illustrate the potential harms of secondary uses of patients' data. This is not to suggest that no socially beneficial purpose was pursued within these case studies. But in all three, I identify some degree of disproportionate, or indirectly discriminatory, interference with Article 8 ECHR. As such, they serve as case studies to argue for a greater *scope* of privacy protection,[29] and for a more expanded model of evaluation and *justification* of secondary uses of patients' data.[30]

It is not necessary to set out, here, every statutory provision that applies to secondary uses of patients' data. But I will provide an overview of three key provisions, to help explain the powers and obligations at play within the case studies explored in my doctoral publications:

1. Section 261 Health and Social Care Act 2012 ('HSCA'), which allows for disclosure of anonymous patient data. This provides a background to Part 1 of this thesis, which focuses on Identification, and when data can be considered anonymous.
2. Section 251 NHS Act 2006 ('NHSA'), which allows the Secretary of State for Health and Social Care ('SoS') to mandate the sharing of identifiable patient information. This provides a background to Part 2 of this thesis, which focuses on Private Life.
3. Article 35 GDPR, which sets out the requirement to conduct a Data Protection Impact Assessment ('DPIA'). This provides a background for Part 3 of this thesis, which focuses on Justification.

Section 261 HSCA: Anonymous Data

Scientific research conducted by academics working outside the NHS is an important part of the day-to-day landscape of secondary uses of patients' data in England. Much of the research that leads to new treatments, or changes in clinical practice, is not conducted by the NHS itself, but by academic researchers who must apply to access patient data for their studies.[31] Assuming these applications are successful, NHS Digital (a data-custodian body

[29] Particularly in the form of a legal presumption of privacy in secondary uses of patients' identifiable data in the 4. Reasonable expectations article **(thesis page 63).**

[30] Particularly in the form of an enhanced Data Protection Impact Assessment, as advocated in the 7.DPIA+ article **(thesis page 77).**

[31] C. Metcalfe, R. M. Martin, S, Noble, et al, 'Low risk research using routinely collected identifiable health information without informed consent: encounters with the Patient Information Advisory Group' (2008) 34 Journal of Medical Ethics 37-40.

which now forms part of NHS England[32]) shares patient information on the basis of s.261 HSCA. There are various subsections that NHS England relies upon for these disclosures. Broadly speaking, it suffices to say that s.261 permits NHS England to disclose identifiable and anonymised patient data for secondary purposes connected with the provision of health and social care, or the promotion of health.[33]

To gain a full sense of the scale and variety of these disclosures, the reader would need to undertake the unenviable task of reviewing NHS England's Data Use Registers,[34] which document the recipients of data shared under s.261 HSCA.[35] I have reviewed some of these Registers: the most recent, as at January 2025, documents 25,089 disclosures of patient data made on an anonymised basis (74% of the total), compared with 8,689 disclosures (26%) on an identifiable basis. This is consistent with previous work I have done, which is not included in this thesis but forms an important part of its background.[36] In December 2016 to May 2018, between 76% and 86% of releases documented in the registers were justified on the basis of s.261, combined with anonymisation of the data.[37]

In Part 1, I will consider the criticisms of anonymisation as a means of safeguarding data subjects' privacy interests. I will also consider how a 'privacy-based' approach might indicate ways of using patients' data without interference with their private life. As such a large proportion of disclosures outside the NHS occur on an anonymised basis, the concept of anonymity warrants further discussion. This Part will consider the anonymous vs identified binary in more depth, and advance a privacy-based account of identity, based on interference with Article 8. As anonymous disclosure is evidently such an important part of the landscape of secondary uses of NHS data, it is in patients' interest to have a coherent account of what 'anonymous' means in the context of **Scientific Research**, and **Tissue Donation.**

Section 251 NHSA: Identifiable Data

Part 2 of this thesis focuses on identifiable patient data, and how these map onto the scope of 'Private Life' for the purposes of Article 8 ECHR. I will argue that all secondary uses of identifiable patient data should engage the right to private life under Article 8. As demonstrated in the context of the **Immigration** and **DeepMind** case studies, however, it is not currently guaranteed that such uses of patients' information will attract a 'reasonable

---

[32] NHS England, 'NHS Digital and NHS England complete merger' (1 February 2023) https://www.england.nhs.uk/2023/02/nhs-digital-and-nhs-england-complete-merger/.
[33] As required by s.261(1A) Health and Social Care Act 2012.
[34] NHS England, 'Data Use Registers' https://digital.nhs.uk/services/data-access-request-service-dars/data-uses-register#data-uses-registers.
[35] A few documented disclosures in the January 2025 Register, for example, were made on other legal bases, such as the Covid-19 Regulations, the Statistics and Registration Service Act 2007, and s.251 NHS Act 2006. But for the sake of this overview it suffices to say that most disclosures for research are made under s.261 Health and Social Care Act 2012.
[36] M. Mourby, J. Doidge, K. H. Jones et al, 'Health Data Linkage for UK Public Interest Research: Key Obstacles and Solutions' (2019) International Journal of Population Data Science 4.1.
[37] Mourby et al, ibid.

expectation of privacy' in English law. This is due— in part— to the statutory definition of 'confidential patient information' under s.251 NHSA, which must include:

> a)information (however recorded) which relates to the physical or mental health or condition of an individual, to the diagnosis of his condition or to his care or treatment, and
>
> (b)information (however recorded) which is to any extent derived, directly or indirectly, from such information[38]

This has been taken, in some judgements, to mean that NHS patient information which does not relate to physical or mental health is not confidential, and thus does not attract a reasonable expectation of privacy.[39] In Part 2 of this thesis, I will challenge this construction of privacy and confidentiality, and suggest that patient data does not need to reveal health-related information (directly or indirectly) to interfere with Article 8. As I will explain, the *scope* of privacy and confidentiality rights in identifiable patient data is important, because it changes how the use of patients' data should be *justified*. I will argue that Article 8 requires more robust, systematic justification of interference with patients' private lives than is routinely conducted when the SoS exercises their powers under s.251 NHSA to require the transmission of identifiable patient information.

Another key provision for this thesis, therefore, is the power vested in the SoS under s.251 NHSA to mandate the transmission of confidential patient information. Section 251 NHSA allows the SoS to mandate this transmission on a time-limited basis: the letters communicating these 'Control of Patient Information Notices'[40] have a stated time-limit for the authorisation in question. However, they are not necessarily accompanied by any evidence that the broader lawfulness of the processing has been evaluated— for example, under data protection law, or the Equality Act 2010. This has implications for Part 3 of this thesis, which focuses on the justification of secondary uses of patients' data.

Article 35 GDPR: Justification

Notices from the SoS under s.251 NHSA can place NHS bodies in a difficult position. A mandate to share patient data from the SoS does not mean that all other legal requirements have been considered. The primary concern of these 'Control of Patient Information' notices, issued under s.251 NHSA, is to ensure that there is no breach of medical

---

[38] Section 251(10) NHSA. This is supplemented by s.251(11), which stipulates that this health-related 'patient information' is 'confidential' when it is identifiable.

[39] *R (on the application of W, X, Y, and Z) v The Secretary of State for Health v The Secretary of State for the Home Department* [2014] EWHC 1532 (admin), [45]-[46]. However, the Court of Appeal disagreed with the High Court's logic, on this conclusion, and suggested that NHS patient data could be confidential even without directly revealing health related information, see [2015] EWCA Civ 1034, note 88, [34]-[40].

[40] Made by the SoS under the Control of Patient Information Regulations 2002.

confidentiality when the information is disclosed.[41] But NHS bodies also have obligations under the Data Protection Act 2018, and under public law in general, which cannot be swept aside by a notice from the SoS.[42] There is no explicit legal mechanism under s.251 NHSA for NHS bodies to decline any aspect of the instructed processing, for example to make their own public interest assessment. Nonetheless, NHS England has been criticised for its lack of independence from the government in complying with requests by the SoS to share patients' information.[43] In the 7.DPIA+ article, I argue that one way to ensure that the use of patients' data is lawful beyond the requirements of confidentiality is to conduct a Data Protection Impact Assessment, under Article 35 GDPR, as enhanced by joint consideration with the Equality Act 2010. This argument is made in the context of the **Covid-19 Case Study**, but a 'DPIA+' would be a useful tool in other contexts.

For example, in the **DeepMind Case Study** in Part 2 of this thesis, I examine the relationship between the Royal Free NHS Foundation Trust ('Royal Free') and Google UK Ltd. In disclosing patient data to Google, the Royal Free acted outside the SoS's regulatory system under s.251 NHSA,[44] using what appeared to be (perhaps in hindsight) broad powers under s.43 NHSA.[45] Despite the statutory basis for this disclosure, which might have been enough to satisfy medical confidentiality, Royal Free's data protection obligations were not fulfilled. The Information Commissioner's Office publicly condemned the transmission of identifiable patient data as constituting multiple breaches of the (then) Data Protection Act 1998.[46] Despite the legal basis within the NHSA, therefore, the disclosure evidently fell short of lawfulness in any holistic sense, that considers multiple areas of law. In particular, the potential for disproportionate impact on certain categories of patient[47] is, I argue, a factor that could have been addressed by a more holistic approach, such as within my 'DPIA+.'

A 'DPIA+' would also have been a useful *ex-ante* tool within the **Immigration Case Study**. Neither the NHSA, nor the HSCA, prompted the SoS or the NHS to consider the totality of

---

[41] Under s.251(2)(c) NHSA, Regulations made under this provision ensure that any processing carried out under the SoS will be lawful '*despite any obligation of confidence*'— i.e., lawful at least according to the requirements of common law confidentiality.

[42] These additional obligations are considered in Part 3 of this thesis, particularly within the proposed 'DPIA+' model.

[43] Health and Social Care Committee, *Memorandum of understanding on data-sharing between NHS Digital and the Home Office* (2017-9, HC 677) 3.

[44] A pointed emphasised by the Court of Appeal in subsequent litigation—possibly because the s.251 NHSA 2006 system would trigger the application of patient opt-outs from data-sharing. See *Andrew Prismall v Google UK Ltd & Ors* [2024] EWCA Civ 1516, [9].

[45] Per an audit conducted by Linklaters LLP. As with the immigration guidance, however, this may or may not have been a retrospective rationalisation of the data-sharing, to align it with a plausible statutory power: Linklaters LLP, 'Audit of the acute kidney injury detection system known as Streams' (17 May 2018, updated 7 June 2018) https://www.royalfree.nhs.uk/application/files/1516/9721/4007/Streams_Report.pdf, page 43.

[46] E. Denham, Letter to Sir David Sloman (3 July 2017) https://ico.org.uk/media/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf.

[47] For example, pregnant patients whose medical abortions were disclosed for the purpose of developing a kidney injury app. See H. Hodson, 'Revealed: Google AI has access to huge haul of NHS patient data' (29 April 2016) https://institutions.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/.

their overlapping evaluative duties when entering a Memorandum of Understanding ('MOU') to share information with the Home Office.[48] But when these disclosures are challenged, they are challenged in their totality, and then all alleged breaches of data protection and anti-discrimination are raised together. For example, the NHS's MOU with the Home Office was challenged on the grounds that:

> it breaches privacy and data protection rights under the Convention and the Charter, discriminates against patients who are subject to immigration control; indirectly discriminates against disabled and female migrants; and that the defendants are in breach of the PSED [i.e. the Public Sector Equality Duty under s.149 Equality Act 2010][49]

These areas of law—data protection, privacy and anti-discrimination—cluster together in multiple actions against public bodies,[50] and I will explore their confluence in Part 3 of this thesis. In my view, these causes of action often arise together because they share a common source: the ECHR. This is the final aspect of the legal landscape for secondary uses of patient data, as explained below in subsection B. I will argue that this common legal underpinning also means that breaches of these overlapping duties may cluster together, and there are thus key pragmatic advantages in considering them together, *ex-ante*.

In sum, however, the broad powers given to the Secretary of State under the NHS Act 2006 are of particular concern from a privacy perspective, as are the general powers of the NHS leadership to make bilateral arrangements to disclose patients' data to software companies. These powers within the legal framework are most in need of checking, in terms of the necessity and proportionality of their exercise for specific programmes of secondary data use. This is why 'top-down' secondary uses of patients' data, such as the **Immigration Case Study,** the **DeepMind Case Study** and the **Covid-19 Case Study,** will feature prominently in Parts 2 and 3 of this thesis, which examine the potential privacy harms of secondary uses, and the potential benefit of Article 8 ECHR as an evaluative tool to prevent/ mitigate these harms.

---

[48] A copy of the (now withdrawn) 2017 MOU is available at: Home Office, Department of Health and NHS Digital, 'Memorandum of Understanding between Health and Social Care Information Centre and the Home Office and the Department of Health' (1 January 2017) https://assets.publishing.service.gov.uk/media/5c4f2103e5274a492e19de96/MoU_between_HSCIC__Home_Office_and_DH.pdf.

[49] Matrix Chambers, 'Migrants' Rights Network granted permission for judicial review of patient data-sharing agreement between NHS Digital and the Home Office' (1 March 2018) https://www.matrixlaw.co.uk/news/migrants-rights-network-granted-permission-legally-challenge-data-sharing-agreement-nhs-digital-home-office/.

[50] See the 7.DPIA+ article for a discussion of *R (MXK, EH, HH, SXB, and ALK) v Secretary of State for the Home Department* [2023] EWHC 1272 (admin), and *R (on the application of Edward Bridges) v The Chief Constable of South Wales Police & Ors* [2020] EWCA Civ 1058.

I will explain the significance of Article 8 for secondary uses of patients' data in the next subsection.

B. Article 8 ECHR as a Fundamental Value

Part 3 of this thesis will consider how Articles 8(2) and 14 ECHR can strengthen patients' protection when their data are re-used for secondary purposes. It will do so by focusing on the requirements of proportionality[51] and non-discrimination.

Beneath the fragmented jigsaw of powers, and the complex system of data protection 'checks' on these powers, lies a standard which, in my view, could help bring the NHS information governance system into balance. The regulation of patients' data in English law is ultimately underpinned by Article 8 ECHR. This is not a novel or controversial view of information law in this jurisdiction, stemming as it does from a commonly held view of the UK's unwritten constitution, particularly after the Human Rights Act 1998.[52] Rights under the ECHR function as standards against which laws passed by parliament, or regulations passed by the SoS, can be judged. To use Article 8 as a tool to evaluate English law is, therefore, not a novel exercise. It is an increasing feature of our legal system that doctrine evolves in response to the influence of ECHR rights as an interpretive benchmark.

English information law, which provides the protections for patients when their personal data are re-used, has evolved in response to Article 8 ECHR. The most obvious example is the new tort of misuse of private information (or 'MOPI'), which evolved out of the law of confidentiality to give effect to Article 8 ECHR outside of pre-existing relationships.[53] Within the law of confidentiality, there is some disagreement as to whether equitable principles, such as the 'conscience' test, should continue to be the predominant benchmark.[54] Nevertheless, there is at least one line of authority under which the law of confidence should be construed via an ECHR-inspired principle of proportionality.[55] The other strand of information law, data protection, is much more explicitly linked to Article 8 ECHR.[56] This is

---

[51] For ease of reference throughout this thesis, the term 'proportionality' is used to connote the assessment of proportionality which should take place to determine whether an interference with (inter alia) Article 8 ECHR is necessary in a democratic society, and represents no greater intrusion into a Convention right than is necessary for the legitimate aim pursued. See, for example, *Handyside v The United Kingdom (*1979-80) 1 EHRR 737.
[52] T. R. Hickman, *Public Law after the Human Rights Act* (Oxford: Hart Publishing, 2010); A. Young, *Democratic Dialogue and the Constitution* (Oxford: Oxford University Press, 2017).
[53] Beginning with the UK House of Lords judgment in *Campbell v MGN Ltd* [2004] UKHL 22; [2004] 2 AC 457, at [14].
[54] This is a position criticised by M. J. Taylor and J. Wilson (note 23) but defended by E. S. Dove in 'Misuse of private information and the common law right of privacy: a new frontier in biomedicine?' in Dove (ed) *Confidentiality, Privacy and Data Protection in Biomedicine* (Abingdon: Routledge, 2024), 194-231, 197.
[55] *London Regional Transport v Mayor of London* [2001] EWCA Civ 1491, [2003] EMLR 4 [58].
[56] The GDPR references, in its first recital, the right to data protection under the EU Charter on Fundamental Rights, which is itself derived from Article 8 ECHR— see 'Explanatory Memorandum, Convention document CHARTE 4473/00' (11 October 2000) https://www.europarl.europa.eu/charter/pdf/04473_en.pdf. Before the GDPR, the Data Protection Directive 95/46 EC (see note 57 below) referenced Article 8 ECHR in its first recital, as this generation of data protection preceded the Charter.

part of the reason why data protection, and the GDPR, are so much the focus of Parts 1 and 3 of this thesis: it is an area of information law which is explicitly connected to Article 8, and thus has significant potential to import its values.

The evolution of these three areas of law has taken place, slowly, since the first EU data protection legislation (the Data Protection Directive[57] in 1995), and the Human Rights Act 1998. My 're-evaluation' is therefore partly observational/analytical, in the sense that I am delineating legal developments that exist outside of my own work. However, there is also a normative element to my doctoral evaluation, at least in the emphasis I am placing on these trends. By accepting and supporting Article 8 ECHR as a standard which both does, and *should*, influence health privacy law, I am aiming to nudge the law further along this existing direction of travel.

This nudge is necessary to improve legal protection of patients' data within NHS information governance. Article 8 ECHR does not feature explicitly in most of the day-to-day regulation of patient data in the NHS. None of the statutory powers I have described above, under which the SoS can mandate large programmes of secondary data use, require the SoS to undertake (for example) a Data Protection Impact Assessment that could capture any human rights implications under Article 8. Even within confidentiality guidance aimed at NHS bodies, the Human Rights Act 1998 is not construed as adding much value, concluding as it does:

> In general, compliance with the Data Protection Act 1998 and the common law of confidentiality will satisfy HRA requirements. However, this is a complex area of law that is open to interpretation by the courts, meaning that specific legal advice should be sought to ensure compliance in the particular circumstances.[58]

The tentative conclusion of this guidance— that compliance with data protection and confidentiality law will satisfy most concerns about Article 8 ECHR— is only correct if these areas of law are understood expansively. Where they are construed narrowly, without the influence of patients' privacy rights, they are very flawed vessels for the protection that Article 8 ECHR should afford UK citizens under the Human Rights Act 1998.

The seven publications included in this thesis support the proposition that Article 8 ECHR should be a key benchmark in the regulation of secondary uses of patients' data (and citizens' data more broadly). Some of these pieces have now been included in this 'publication format'

---

[57] 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, OJ 1995 L 281/31 (hereafter cited as 'the Data Protection Directive').

[58] NHS Digital, 'A guide to confidentiality in health and social care: references. Section 4: Human Rights Act provisions' (13 January 2022), available from: https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/codes-of-practice-for-handling-information-in-health-and-care/a-guide-to-confidentiality-in-health-and-social-care/hscic-guide-to-confidentiality-references/section-4

thesis, with Article 8 ECHR forming a central thread that ties together the 'privacy-based' approach of this thesis. The next section explains the publication format in more detail.

IV. Publication Format

This thesis is in 'publication format,' meaning it contains seven journal articles and book chapters which have already been published, or which have been submitted for publication. For ease of reference, these pieces have been incorporated in their original font and format of the author accepted manuscript, so that they can be visually distinguished from the integrative text of the thesis with relative ease. I have also retained the original referencing style and numbering, so that each publication is self-contained for the purposes of cross-referencing (again, to make the overall manuscript less unwieldy to navigate). Each publication of this thesis is self-contained in its footnoting. For example, wherever a footnote within a publication says 'ibid note 21', this is a reference to a footnote within that publication.

All these publications were written with the ultimate ambition of inclusion within a PhD thesis, although the earlier papers come from collaborative projects on which it was expected practice to be co-authors on each other's papers. To give more background on the origin of each publication, and explanation as to why each is sufficiently authored by me to justify inclusion in my PhD thesis, each Part of this thesis has a 'Synopsis' where I explain in more detail how each paper was written, as well as its connection to the central thread of this thesis. In short, I am the lead author of all co-authored papers. Unless explicitly specified in a Synopsis to the Part, I wrote all the text of the joint papers, with the acknowledged benefit of the comments and feedback of my co-authors.

Although all the pieces were written with the ultimate ambition of a doctoral thesis, the later publications were written when this thesis had taken on clearer shape, and as such are more neatly aligned with its aims and scope. For example, the first article in Part1 focuses on secondary uses of public sector data in general— rather than NHS patient data in particular. This does not, in my view, detract from the broader study of identity as interference with private life, which is the purpose of Part 1. I have reflected carefully on which publications to include in this thesis, and will explain why I believe each assists the development of my argument.

V. Methodology

I have used a legal doctrinal methodology both in the preparation of the publications which serve as publications in this thesis, and in writing the integrative text that ties them together. It is the method best suited to my thesis' research question: 'can a privacy-based approach help regulate secondary uses of patients' data?' As explained above, this research question re-interprets the existing legal framework for secondary uses of patients' data. As I am using a legal doctrinal conception of privacy in my 'approach', what follows in this thesis is naturally a legal doctrinal discussion.

My overarching aim is to re-align the laws governing NHS patient data with a fuller understanding of the right to privacy under Article 8 ECHR. Such a re-evaluation of the law falls well within the aims and purpose of doctrinal legal study. Although legal doctrinal research can be a difficult, intuitive method to define, it is generally understood as a search for:

> a systematic exposition of the principles, rules and concepts governing a particular legal field or institution and analyses the relationship between these principles, rules and concepts with a view to solving unclarities and gaps in the existing law.[59]

This description broadly aligns with my own doctrinal aim, with the nuance that my doctrinal approach goes beyond the merely 'expository.' Exposition is indeed an initial stage of my analysis. For example, the links between Article 8 ECHR and the English law of confidentiality, the MOPI tort and data protection law are not my own inventions: they are pre-existing aspects of the law which my research (and the research of others[60]) has confirmed. However, the second stage of my work is normative. Having identified this link between Article 8 and English health privacy law, I argue that it should be strengthened, through re-interpretation of these areas of law in greater alignment with Article 8.

This normative dimension falls well within the predominant understanding of doctrinal legal scholarship. Indeed, purely expository understandings of doctrinal methodology have been characterised as a minority position.[61] Rather than objectively revealing the law, a doctrinal legal scholar is generally understood as identifying and expounding a case for its 'better' interpretation. A key voice in this discussion is that of Martha Minow, who sees a 'doctrinal restatement' of the law as an attempt to:

> a. Organize and reorganize case law into coherent elements, categories, and concepts;
> b. Acknowledge distinction between settled and emerging law;
> c. Identify difference between majority and "preferred" or "better" practice–- ideally with some explanation for the criteria to be used.[62]

The key objective of this thesis, per Minow's schema, is the identification of some 'better' approach to regulating patients' rights in secondary uses of health data. On what basis can I contend that my approach is 'better' than the current one? Here, I am borrowing lightly from

---

[59] R. van Gestel, H. Micklitz and E. Rubin, 'Should Doctrinal Legal Scholarship Be Abandoned?' in van Gestel, Micklitz and Rubin (eds) *Rethinking Legal Scholarship: A Transatlantic Dialogue* (Cambridge: Cambridge University Press, 2017), 205–398.
[60] For example, M. Taylor and J. Wilson on the role of Article 8 ECHR in the English law of confidentiality. See note 23.
[61] A. R. Mackor, 'Explanatory Non-Normative Legal Doctrine, Taking the Distinction between Theoretical and Practical Reason Seriously' in Van Hoeke (ed) *Methodologies of Legal Research : Which Kind of Method for What Kind of Discipline?* (2013, Bloomsbury Publishing: London), 45-70.
[62] M. Minow, 'Archetypal Legal Scholarship: A Field Guide' (2013) 63 Journal of Legal Education 1, 65-69.

critical legal studies– not as a methodology in and of itself, but as an 'attitude,'[63] that centres people who may otherwise become marginalised within legal orthodoxy.

Which patients do current confidentiality, privacy and data protection laws serve least well? In posing this question in the context of privacy rights, Beitz's words are helpful:

> human rights need not be interpreted as deriving their authority from a single, more basic value or interest such as those of human dignity, personhood, or membership. The reasons we have to care about them vary with the content of the right in question and the nature of our relationship, if any, with various classes of potential victims of abuse.[64]

Throughout this thesis, I have tried to maintain focus on the individuals most affected within my case studies of secondary uses of patient data. These people would include overseas visitors vulnerable to immigration consequences from the disclosure of their NHS data, and people without the 'data-awareness' to opt in or out of information sharing initiatives.[65]

Fundamentally, this thesis is more concerned with the impact on the private lives of these people, than with any abstract conception of autonomy, dignity, or any other value which could be seen as underpinning the right to privacy. My approach thus seeks to benefit these people, and not this concept. I will take interference with Article 8(1) ECHR as the key benchmark in exploring 'Identification' and 'Private Life.' This focus on interference enables considerations of impact on data subjects. It means the key initial question, on a practical level, should be *how might this secondary use of information affect a patient's private life?* If it is established that the data are personal and private,[66] the subsequent question becomes *what kind of patient might be most affected, and is this to a degree that is this impact necessary to the aim pursed?* This goes to the question of justification under Article 8(2).

---

[63] P. Minkkinen 'Critical legal "method" as attitude' in Watkins (ed) *Research Methods in Law* (London: Routledge, 2017), 146-169, 148.
[64] C. R. Beitz, *The Idea of Human Rights* (Oxford: Oxford University Press, 2009).
[65] See Delacroix and Lawrence, note 77.
[66] Under my 'privacy-based approach,' I will use interference with Article 8(1) as the essential touchstone for defining personal data (in Part 1) *and* private information (in Part 2).

VI. Contribution

A. Practical contribution

The problems I am addressing have been outlined above: the disparate powers to share patient data for secondary uses in England, the controversies that have dogged their recent exercise, and the lack of clarity around the protection of patients' interests. I am clearly not the first person to respond to these problems in the secondary uses of patient data. But my contribution lies in re-grounding NHS information governance in the right to privacy, and (in particular) a version of this right that looks beyond the value it places on individual autonomy. In addition to the theoretical importance of re-shifting the central point of the debates from data protection back to privacy, as will be explained in the next subsection, the contribution of this thesis is also of practical importance. It provides a re-interpretation of the legal framework that could help make future secondary uses of patients' data more legally sound, across multiple areas of law. My account of Article 8 ECHR is supplemented by Article 14 ECHR, which states:

> The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.

In other words, Article 8 has a built-in anti-discrimination principle, due to the structure of the ECHR as a whole. This anti-discrimination principle prompts the question of whether any type of patient would be disproportionately affected by secondary uses of health data. This re-emphasis of the legal basis for considering patients as groups, and not just as individuals, is thus a secondary contribution of this thesis. It is a consideration which comes to the fore in the judgments which have become my de facto case studies.

Practical Contribution: Immigration Case Study

To explain the importance of my thesis's re-centring of Article 8 (and 14) ECHR, it is helpful to the **Immigration Case-Study.**

When a 2011 change in Immigration Rules introduced a sanction for non-citizen NHS debtors,[67] NHS Digital shared the data of thousands of patients with the Home Office.[68] After

[67] [2014] EWHC 1532 (admin), note 39, [1].
[68] The Health and Social Care Committee report (note 43, report page 6) states that there were 6,774 tracing immigration requests in 2015-2016 alone.

years of litigation,[69] parliamentary criticism[70] and public consultation[71] around these data-sharing arrangements, they were eventually abandoned in 2018, with the Home Office agreeing to focus its monitoring on criminal offenders,[72] rather than patients who had been unable to pay for their healthcare. Before this programme of data sharing began, upstream assessment of its impact on different demographic groups (e.g. pregnant women, and people with disabilities) could have revealed its disproportionate impact on patients less able to avoid seeking medical care.

In particular, a combined assessment informed by Articles 8-14 ECHR should have highlighted the risk of discriminatory impact in this use of patients' information, which should have informed a consideration of the policy's scope and justification.

Practical Contribution: DeepMind Case Study

Likewise, another key case study for this thesis is the Royal Free's disclosure of patient data to Google **DeepMind**. As I argue in the 7.DPIA+ article, a group-based analysis of privacy risk could have helped identify the types of patient who could have been disproportionately affected by the disclosure. This could have helped filter out patients for whom the risks of disclosure could significantly outweigh the benefits of developing a kidney injury app. Had these patients been excluded, this might, in turn, have helped avoid the Information Commissioner's finding that the data-sharing breached the proportionality and data minimisation requirements of data protection law.[73] The NHS, in this case, was not assisted by a simple construction of the private law, which directs attention to an 'objective' perspective of a reasonable expectation of privacy.[74] The added-value of the public law perspective, imported by Articles 8 & 14 ECHR, is the prompt to think beyond an imagined 'objective, reasonable' person, and instead consider the plurality of people who might be disproportionately impacted, such as patients with gender or disability related conditions (e.g. abortions, or positive HIV status[75]).

---

[69] As well as the Court of Appeal judgment in [2015] EWCA Civ 1034 (note 88, below) there was also a challenge under the Equality Act 2010 in *R (MXK, EH, HH, SXB, and ALK) v Secretary of State for the Home Department* [2023] EWHC 1272 (admin).
[70] Health and Social Care Committee report, note 43. See also Letter from Dr Sarah Wollaston MP to NHS Digital (29 January 2018): https://publications.parliament.uk/pa/cm201719/cmselect/cmhealth/Correspondence/Wilkinson-2018-01-29.pdf.
[71] Public Health England, 'Data-sharing MoU between NHS Digital and Home Office: call for evidence' (26 March 2018) https://www.gov.uk/government/calls-for-evidence/data-sharing-mou-between-nhs-digital-and-home-office-call-for-evidence#FULL-PUBLICATION-UPDATE-HISTORY.
[72] Home Office, 'Home Office in the media: 10 May 2018' (10 May 2018): https://homeofficemedia.blog.gov.uk/2018/05/10/home-office-in-the-media-10-may-2018.
[73] Letter from E. Denham, note 46.
[74] As confirmed by the UK Supreme Court in *Bloomberg LP v ZXC* [2022] UKSC 5; [2022] AC 1158.
[75] See Hodson, note 47.

B. Theoretical contribution

My working definition of privacy, within my 'privacy-based approach,' has two key features:

1) It applies the core language and concepts of Article 8 ECHR itself to the wide array of legislation and common law that governs NHS patient data. As such, it represents a 're-grounding' of the relevant law in this fundamental, constitutional right.
2) It understands the values inherent within Article 8 ECHR in a way that looks beyond the value of individual choice, consent and autonomy.

As a preliminary point, the fact that I am applying my approach to consider exclusively legal solutions is, in itself, a distinguishing factor of my doctoral work. Many of the authors concerned with privacy in secondary uses of health data (or secondary uses within Big Data, more broadly) have looked beyond existing law to remedy perceived deficiencies in the status quo. They have proposed new law,[76] or new governance models drawing on existing law to regulate data,[77] as well as new evaluative models for automated analytics,[78] and (more broadly) new models of governance that would cut across multiple actors in the health data ecosystem.[79]

My approach in this thesis does not require any new organisations, re-organisation of existing bodies, or new legislation. It requires only a shift in perspective in how we understand the *scope* of privacy rights (Parts 1 and 2 of this thesis), and how interference with these rights should be *justified* (Part 3 of this thesis) This does not make my aims any less ambitious, however, as our understanding of health data privacy is foundational for all subsequent

---

[76] S. Wachter and B. Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review, 494-620.

[77] A. McMahon, A. Buyx, and B. Prainsack, 'Big Data Governance Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond.' (2020) 28 Medical law review 1, 155–182.; S. Delacroix and N. D. Lawrence, 'Bottom-up Data Trusts: Disturbing the "One Size Fits All" Approach to Data Governance' (2019) 9 International data privacy law 4, 236–252; B. J. Evans, 'Big Data and Individual Autonomy in a Crowd' in Cohen, Fernandez, Vayena and Gasser (eds) *Big Data, Health Law, and Bioethics* (Cambridge: Cambridge University Press, 2018), 19-29.

[78] M. E. Kaminski and G. Malgieri, 'Algorithmic impact assessments under the GDPR: producing multi-layered explanations' (2021) 11 International Data Privacy Law 2, 125–144; Ada Lovelace Institute, 'Algorithmic impact assessment: user guide' (8 February 2022) https://www.adalovelaceinstitute.org/resource/aia-user-guide/; A. Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34 Computer Law & Security Review 4, 754–772; A. Mantelero and M. Esposito, 'An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems' (2020) 41 Computer Law & Security Review 105561; A. Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI* (The Hague: Springer, 2022).

[79] Kaye et al, note 24. Note, however Iusmen and Boswell's scepticism over the incorporation of this model into NHS England's governance (note 27); E. Vayena and A. Blasimme, 'Health Research with Big Data: Time for Systemic Oversight.' (2018) 46 The Journal of Law, Medicine & Ethics 1, 119-129.

solutions. Much of the existing literature suggests that current laws are insufficient to regulate secondary uses of data in a Big Data context. In particular, many scholars have focused on the insufficiency of individual choice or consent as a means to regulate large-scale secondary uses of health data.[80] It is implicitly assumed that these individualistic provisions are the only, or at least the most significant, aspects of privacy law for the regulation of health data.

I differ from the above-cited authors at the fundamental level of this premise. Where current privacy laws offer people insufficient protection, I suggest this can be addressed through re-interpretation, rather than large-scale reform. With renewed connection to its constitutional origin in Article 8 ECHR, other values— such as proportionality, and the Article 14 principle of anti-discrimination— come to the fore as a vital supplement to that of individual autonomy.

C. Article 8 ECHR— Beyond Autonomy

As regards the second point of this thesis's theoretical contribution—privacy beyond autonomy—my approach to privacy is in alignment with that taken by Daniel Solove when he observed:

> With each sign of failure of privacy self-management, however, the typical response by policy-makers, scholars, and others is to call for more and improved privacy self-management. In this Article, I argue that in order to advance, privacy law and policy must face the problems with privacy self-management and start forging a new direction.[81]

I agree with this call for a new direction. However, for the challenges facing secondary uses of patient data in England, I believe this direction can be taken within the confines of existing law. As long as English law is (re)interpreted in alignment with an expansive understanding of Article 8 ECHR, there is potential for confidentiality and data protection laws to extend far beyond the 'self-management' model of consent and opt-outs. For example, they would also encompass the NHS's justificatory and anti-discrimination obligations under public law.

These new readings of confidentiality and data protection law *could* lead to new methods of justifying or evaluating secondary uses of data, such as the DPIA+ model I outline in Part 3 of this thesis. But such proposals are, ultimately, still consolidations of existing law. The main contribution of my thesis is thus an initial, doctrinal foundation for these more practical

---

[80] See McMahon et al, Delacroix and Lawrence, and B. J. Evans (note 77). See also M. Doerr and S. Meeder, 'Big Health Data Research and Group Harm: The Scope of IRB Review' (2022) 44 Ethics and Human Research 4, 34-38; 'Big Data: Destroyer of Informed Consent' (2019) 18 Yale Journal of Health, Policy, Law and Ethics 3, 27-54; J. E. Cohen, 'Turning Privacy Inside Out' (2019) 20 Theoretical Inquiries in Law 1, 1-22.

[81] D. Solove, 'Privacy self-management and the consent dilemma' (2013) 126 Harvard Law Review 7, 1880-1903, 1881.

suggestions; drawing together Article 8 ECHR with the piecemeal framework for patient data into a more expansive understanding of privacy rights in patient data. The value of this prior reflection on the fundamental requirements of the law is that, when new models for regulating secondary uses of health data are proposed, these solutions can reflect values beyond the informational 'self-management' understanding of autonomy-based privacy. A more expansive understanding of Article 8 in this context imports proportionality, public interest and non-discrimination. It provides a legal doctrinal underpinning for consideration of the types of patient who may be particularly affected by secondary uses of their data, due to their race, gender, nationality, age or disability.

This expansive understanding of Article 8 ECHR leads me to my second, more nuanced, claim to originality. Even among the smaller number of authors who have focused on existing law in their studies of health data privacy—rather than proposing solutions outside the law--their characterisation of Article 8 ECHR can focus more on its protection of autonomy, consent and individual choice.[82] This reflects broader, international perspectives on human rights in health data. For example, a report on AI in Healthcare from the World Health Organization states:

> From a human rights perspective, an individual should always control his or her personal data. Individuals' right to their own data is grounded in concepts that are related to but distinct from ownership, including control, agency, privacy, autonomy and human dignity.[83]

This emphasis on autonomy as an essential aspect of human rights in health data is understandable. As well as reflecting a number of national legal traditions,[84] scholarship in this area can be influenced by bio-ethical literature, which has recently placed significant focus on autonomy as a 'good' to be preserved in healthcare.[85]

Autonomy is, undoubtedly, important within healthcare law more broadly. Particularly in the context of bodily privacy, where patients must provide informed consent to prevent a physical intervention from constituting an assault.[86] But, in the context of secondary uses of health data, autonomy, consent and individual choice are not necessarily the most useful or pertinent values at play. From my review of the case studies in this thesis, I believe that other aspects of the right to private life under Article 8 ECHR— such as proportionality and the

---

[82] For example, V. Chico, *Genomic Negligence: An Interest in Autonomy as the Basis for Novel Negligence Claims Generated by Genetic Technology* (Abingdon: Routledge-Cavendish, 2011).
[83] World Health Organization, 'Ethics and governance of artificial intelligence for health' (28 June 2021) https://www.who.int/publications/i/item/9789240029200.
[84] I will touch on Germany as a counter-example to the UK in the 6.EHDS chapter.
[85] R. Gillon, 'Ethics Needs Principles—Four Can Encompass the Rest—and Respect for Autonomy Should Be "First among Equals"' (2003) 29 Journal of Medical Ethics 5, 307-312.
[86] T.T. Arvind and A. McMahon. 'Responsiveness and the Role of Rights in Medical Law: Lessons from *Montgomery*' (2020) 28 Medical Law Review 3, 445–477.

importance of non-discrimination—[87] have a role to play in regulating secondary uses of patients' data. I will briefly outline why, with particular reference to the **Immigration Case Study:**

Autonomy in the Immigration Case Study

When the NHS transmitted information to the Home Office for immigration purposes, there was a much-contested question as to whether this was compatible with the NHS's duty of confidentiality, as underpinned by Article 8 ECHR. In a paper which has been very important for my thesis, Mark Taylor and James Wilson argued that, on the facts of *R (W, X, Y & Z),*[88] the NHS had fallen short of its duty of confidentiality in its transmission of patient data to the Home Office.[89] I agree with this conclusion, but differ as to *why* I suggest confidentiality law was breached. At page 450 of their article, the authors state:

> It seems to us that the circumstances in which one could be confident that the use of confidential information is consistent with a respect for human autonomy and dignity are likely, minimally, to involve an individual having accessible opportunity to signal preferences in relation to uses they can freely choose to accept or reject: opt outs are available to those who wish to object.

In this specific instance, I do not agree that respect for human autonomy through the availability of opt-outs is the key value within Article 8 that was violated by the disclosure of patient data. To explain why, it is worth considering the scenario practically. If the non-citizen patients were offered the choice to opt out of disclosure of their data to the Home Office, what patient would not exercise it? What patient *would* choose to have their information disclosed to the Home Office, for the purpose of experiencing immigration sanctions? This is not a policy where the choices and interests of individuals can fairly serve as a benchmark, because their interests are so clearly stacked against disclosure. When there is no ambiguity around the interest of the individual, the exercise of that person's autonomy has a limited normative role to play. A more meaningful question, here, is whether the disclosure is necessary in a democratic society— per Article 8(2) ECHR. This would involve weighing the potential impact on these individuals against the public interest in these (non-criminal) disclosures, and indeed the public interest in providing adequate healthcare to non-ordinarily resident populations.[90] It should also involve considerations of whether any particular patient group (such as pregnant women) would be disproportionately affected by the disclosure, and any ultimate immigration sanction.[91]

---

[87] As emphasised by Corina Heri, note 2.
[88] *R (on the application of W, X, Y and Z) v Secretary of State for Health and Secretary of State for the Home Department, and the British Medical Association* [2015] EWCA Civ 1034; [2014] 5 WLUK 528.
[89] Note 23.
[90] A point raised by the parliamentary Health and Social Care Committee, note 43.
[91] As was found to be the case in *R (MXK, EH, HH, SXB, and ALK) v Secretary of State for the Home Department* [2023] EWHC 1272 (admin).

The **Immigration Case Study** illustrates why I use this thesis to advocate for an expansive scope of Article 8 ECHR in secondary uses of health data. I mean 'scope' in both the sense of the *circumstances* in which Article 8 should be considered engaged, but also the scope of *values* the right protects: individual autonomy, but also proportionality, public interest and freedom from discrimination.

D. A Note on 'Approach'

The above analysis should, I hope, provide a thorough explanation of what I mean in by the term 'privacy-based' within this thesis. While it is not as loaded a term, I should also explain what I mean by 'approach,' when I explore a 'privacy-based approach.'

I use the word 'approach' to mean an emphasis on a particular value, or set of values, within the decision-making of multiple actors. This broadly aligns with what Vayena and Blasimme[92] call a 'systemic orientation.' They describe how the health data ecosystem is made up of a complex network of actors and technologies. When they advocate for 'systemic oversight', they are (in my view) essentially promoting certain values to be stressed within the decisions and actions of that network. This is essentially what I envisage, but I prefer the plainer term 'approach.'

This subsection has explained how my thesis contributes, overall, to the governance of secondary uses of patient data in the NHS. I will develop an approach based on 'privacy,' understood expansively via an account of Article 8 ECHR that is grounded in the impact on patients as individuals and groups. This overarching contribution is spread across three sections, which each deliver a different aspect of this contribution:

1) Identity as interference with Article 8(1): an understanding of 'identification' grounded in the potential for interference with a patient's private life.
2) Private life as interference with Article 8(1): rejecting the idea that only 'medical' information is private or confidential, and advancing the argument that all identifiable patient information held by the NHS should be considered both private and confidential when used for secondary purposes.
3) Justification through Articles 8 and 14: centring both the established understanding of necessity, lawfulness and proportionality under Article 8(2), but also the question of disparate impact across different patient groups, which could engage the non-discrimination requirements of Article 14.

The next subsection explains this structure in more detail.

---

[92] E. Vayena and A. Blasimme, note 79.

VII. Structure

The three Parts of this thesis mirror the structure of Article 8 ECHR. The text of Article 8 reads as follows:

> 1 Everyone has the right to respect for his private and family life, his home and his correspondence.
>
> 2 There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

I have used Strasbourg and UK caselaw to distil this text into three core aspects of Article 8:

1. Identification,
2. Private life
3. Justification.

As explained above, in section I of this Introduction, each Part of this thesis re-evaluates one of these concepts as they have manifested in the law governing secondary uses of patients' data. The ultimate aim is to develop a clarified understanding of each concept, which can help regulate secondary uses of patients' data.

Part 1 of this thesis begins with Identification, because this is a preliminary issue for the question of the *scope* of Article 8. It is a fair summary of the law to say that Article 8 is deemed to apply when a person is identified, or at least identifiable.[93] Conversely, when an individual is *not* identifiable from a set of data, the legal authority supports the proposition that Article 8 is not engaged.[94] This brings Article 8 itself in line with its offshoot branches of law: information is not deemed to engage the duty of confidence[95] or data protection law[96] if individuals cannot be identified from it. Therefore, a preliminary question in the scope of Article 8(1) is whether individuals can be identified from the information. If a public authority misunderstands the personal vs anonymous nature of the information in question, their subsequent information governance will be fundamentally flawed.[97]

---

[93] For example, *Vicent del Campo v Spain* (Application no. 25527/13) [2018] ECHR 909.
[94] *South Lanarkshire Council v The Scottish Information Commissioner* [2013] UKSC 55 [26].
[95] *R (Source Informatics Ltd) v Department of Health* [2000] All ER 786.
[96] Under Recital 26 GDPR although this was also the case under the preceding Data Protection Directive.
[97] For example, South Wales police failed to adequately grapple with the application of Article 8 ECHR to the data collected by facial recognition software in *R (Bridges) v The Chief Constable for South Wales* [2020] EWCA Civ 1058, rendering their subsequent Data Protection Impact Assessment inadequate, [152-153].

The question of identification is therefore a key initial issue to the scope of Article 8. In Part 1 of this thesis, I address the concept through a 'privacy-based approach,' linking it to the potential for interference with private and family life, within the meaning of Article 8(1) ECHR.

A. Part 1: Identification

Part 1 contains three publications that build up a privacy-based approach to identification. Over the course of these publications, I establish why anonymous information— which does not impact the privacy of the original data subjects— differs from identifying information, such as a profile. In the latter case, even where an individual is not 'conventionally' identified in the sense of having their face or name revealed, their individual characteristics can still be scrutinised to the extent that constitutes an inherent interference with private life, and therefore an identification. Although these publications were written over a period of five years, they influenced each other and culminated in a conception of identity as based on interference with private life, which provides a benchmark for the NHS in its use of anonymised patient data for secondary uses. The argument of this Part can be summarised thus: that privacy can help form a conception of identity that permits socially beneficial uses of patient data, without causing harm to the private life of individual patients.

A more sceptical approach to secondary uses of patient data, encompassing more of the privacy-related harms to patients in the **Immigration Case Study** and **DeepMind Case Study**, is then taken in Part 2, which considers the scope of private life in the context of patients' identifiable data.

B. Part 2: Private Life

Part 2 considers a second aspect of Article 8(1): whether the information in question relates to a person's '*private life,*' and thus engages the material scope of the right to privacy.

Like 'identity,' the idea of 'private life' for the purposes of Article 8(1) can become associated with types of information, rather than impact on private life. In particular, I have explained above how NHS guidance draws a hard distinction between health-related and demographic data for the purposes of confidentiality.[98] This sits strangely with the idea of private and family life in Article 8, which in its plain English sense would surely encompass a person's private residence. Details such as name and address, while not health-related, can still have an impact on an individual's private life, in the wrong hands. If the law of confidence should be understood as informed by Article 8, then this information should also be understood to have been imparted with the necessary degree of trust and confidence.

[98] NHS Digital, note 12.

This argument is made in Part 2 of this thesis, in the context of the **Immigration Case Study**. This Part only contains one article (the 4.reasonable expectations article), but it develops an important argument for this thesis: that confidentiality and 'misuse of private information' doctrine can become disconnected from the core purpose of Article 8 ECHR, from which they derive:[99] to protect private, family life. I argue for a legal presumption that identifiable patient information is private, within the meaning of Article 8(1) ECHR, given the general likelihood that it could be used in a way that affects patients' private and family life.

This Part considers the arguments for treating all personal data held by the NHS about its patients as private data. As such, it has a greater focus on the top-down harms, which have prompted the European Court of Human Rights to say that the mere storage of personal data by the State should engage Article 8.[100] The case studies which are the focus of Part 2 (the **Immigration** and **DeepMind Case Studies**) highlight the potential for interference with people's private lives.

C. Part 3: Justification

Having established the case for an expanded scope of Article 8(1) (albeit with a clarified understanding of Identification), Part 3 then considers its implications. If we assume, by default, that identifiable patient data are private, then justification of secondary uses under Article 8(2) is also necessary by default. Part 3 therefore considers the added value of justification under Articles 8(2) and 14, as a safeguard in the governance of NHS patient data.

Like Part 1, Part 3 comprises two journal articles and one book chapter. Part 3 is concerned with the evaluation and justification of secondary uses of patients' data. It therefore covers case studies which illustrate the potential benefits of secondary uses (particularly the **Scientific Research Case Study**) but also the potential harms, particularly within the **Covid-19 Case Study.** It is important that these case studies can reflect both the benefits and the harms of secondary uses of patients' data, as this reflects weighing exercise between Article 8 and the Article 10 right to freedom of expression and information (which is discussed in the 5.academic governance article.

Part 1 of this thesis begins on the next page.

[99] The law of confidence precedes the European Convention on Human Rights, but since the Human Rights Act 1998 the personal privacy interests which were historically protected by equity have become closely connected with the scope of Article 8 ECHR. This is explained in more detail in the 4.reasonable expectations article.
[100] *S and Marper v The United Kingdom* (Introduction, note 4).

Part 1 Synopsis: Article 8(1) and Identity

1.1 What is 'Identity?'

A 'privacy-based approach' can help regulate secondary uses of patients' data in multiple ways. The first part of this thesis explores a key initial benefit: a re-evaluation of what it means to be 'identified' by personal information. As per the working definition of 'privacy' in my thesis, my approach grounds the concept of identity in the potential for interference with Article 8(1) ECHR.

What a personal 'identity' means is a question broader, and more multidisciplinary, than this thesis can finally resolve. But it is, nonetheless, a question which must be addressed, at least within the context of secondary uses of patients' data. This is because it is a preliminary question that determines when data protection, confidentiality and privacy laws apply to the processing of patients' data. These laws apply when a person is, or can be, 'identified'— but do not contain conclusive definitions of what kinds of 'identity' should be revealed for this information to constitute an identification.

The concept of identity has shaped the material scope of data protection law since the 1990's, with the introduction of the EU Data Protective Directive 95/46. Recital 26 of this Directive stated:

> Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable.

Since then, guidance such as the Article 29 Working Party's 'Opinion 05/2014 on Anonymisation Techniques' (the 'A29WP guidance') have focused on 'means' of identification— i.e., ways in which a person's identity could be uncovered within information. It is therefore unsurprising that addressing identity, such as by making individuals less or no longer identifiable, has been considered a promising way to minimise the privacy impact while using data, but the *meaning* of identity has not received a commensurate level of attention. I suggest that this conceptual neglect is part of the reason for continuing uncertainty about the scope of data protection.

The uncertainty over the concepts of personal vs anonymous data under the GDPR not only affected organisations that process personal data,[101] but also caused disagreement at the higher level of EU institutions themselves. As recently as 2023, the General Court of the European

[101] Termed 'data controllers' under GDPR Article 4.

Union overruled the European Data Protection Supervisor[102] on its understanding of personal data, and clarified when data could be anonymous for a third party.[103] Only in January 2025— more than six years after the GDPR came into force— did the European Data Protection Board ('EDPB') adopt guidelines on pseudonymisation under the GDPR.[104]

The EDPB's pseudonymisation guidelines were published towards the end of the period in which this thesis was written, but they do (inter alia) confirm one of the main contributions of this Part 1: that pseudonymised data which are personal data for one party can be anonymous for another organisation, as long as broader conditions of anonymity are met.[105] As technical as this distinction may seem, it goes to the core of what (I argue) makes data 'personal', in the sense of revealing identity. If pseudonymised data do not (in the hands of a particular party) give rise to a reasonable likelihood of interference with private and family life,[106] they do not need to reveal identity, and thus do not need to be conceived of as personal data. This preserves the category of 'anonymous' data, and with it the incentive for organisations using medical data for secondary purposes to do so in a way which does not interfere with patients' right to private and family life.

The central argument of this Part 1 is therefore that a 'privacy-based' approach, grounded in interference with Article 8(1), supports a contextual approach to the question of when patients can be 'identified' by their data. Ultimately, this means that medical records (which cannot be deleted at source, for obvious reasons[107]) can nonetheless be re-used for secondary purposes without interfering with patient's privacy, as defined in relation to their informational identity. The initial contribution of my 'privacy-based approach' is, therefore, the use of identity as a legal benchmark to clarify the scope of personal data in patients' records.

1.2 Identity as a legal benchmark

Identification can be a vague concept, both within the GDPR and beyond. The GDPR can be frustratingly inconclusive in making multiple references to the concept of 'identification,' without defining it. I expand the indeterminate nature of identity in the 3.profiling chapter, in which I cite a passage from Clare Sullivan:

> Identity has traditionally been a nebulous notion and in referring to 'identity' without defining it, much of the legal literature in this area lacks precision. It gives the impression that 'identity is identity' whereas the constitution, function and nature of identity depends on context.[108]

[102] The regulatory agency which oversees the data protection compliance of EU bodies.
[103] Case T- 557/20 *Single Resolution Board v European Data Protection Supervisor* [2024] 1 C.M.L.R. 46.
[104] European Data Protection Board, *Guidelines 01/2025 on Pseudonymisation* (Brussels, 16 January 2025).
[105] Ibid, 10.
[106] This potential interference being the core of how I define an 'identification.'
[107] As explored by El Emam and others (see note 129 below).
[108] C. Sullivan, *Digital Identity: An Emergent Legal Concept* (Adelaide: University of Adelaide Press, 2011), 6.

The core contribution of this Part 1 of my thesis is to attempt a more coherent definition of 'identity' in the context of the laws which apply to secondary uses of patients' data in England. The GDPR is a key starting point for a definition of identification, as it offers the main statutory definition of this concept. Although the UK has left the EU, the GDPR has been retained as 'the UK GDPR,' and its core concepts remain relevant for the information law governing secondary uses of patients' data in England. The specific terminology of the Data Protection Act 2018 is outlined in the 3.profiling chapter. Unless explicitly specified, however, references to the GDPR in this thesis refer simultaneously to the EU and UK version of the Regulation, as there is no substantive difference between the two in the context of the present discussion.

The benchmark of identification, as a preliminary threshold for the application of law, has spread beyond data protection to other areas of English doctrine which protect personal information. For example, in 1999, the Court of Appeal confirmed in *R (Source Informatics)* that confidentiality laws would not be breached by the disclosure of anonymised health information,[109] as the patient's only interest at stake was the protection of their privacy, which was safeguarded by the anonymisation. The Court referred to recital 26 of the (then) Data Protection Directive,[110] and the underlying right to privacy under Article 8 ECHR. The clear implication of the Court's reasoning in this judgment is that the right to privacy now underpins what has been (historically) the equitable duty of confidence.

In Part 2 of this thesis, I will examine the evolution of the tort of misuse of private information from the equitable duty of confidence, and how this is also underpinned by the right to privacy under Article 8 ECHR. For the sake of Part 1, however, it is sufficient to say that identification is a key initial trigger for both these areas of law. The precedent in *Source Informatics* still holds for the medical duty of confidence when patients' data are used for secondary purposes.[111]

The initial threshold test for the misuse of private information (or 'MOPI') tort is whether the individual in question had a reasonable expectation of privacy.[112] The caselaw does not explicitly state that an individual must be identified to have a reasonable expectation of privacy, but it was strongly implied in the *Source Informatics* case that there is no Article 8 interest in anonymised information,[113] meaning that the foundational principle underpinning the MOPI tort is not engaged. An individual is only deemed to 'reasonably expect' privacy in information which can, in the first place, be described as private. There will be no 'reasonable expectation of privacy' in information which has been made public by the data subject, for example.[114]

---

[109] *R v Department of Health, ex parte Source Informatics Ltd* (No 1) [1999] EWCA Civ 3011, [2000] 2 WLR 940.
[110] See Introduction, note 57.
[111] See Taylor and Wilson, Introduction note 23.
[112] *Bloomberg LP v ZXC* [2022] UKSC 5; [2022] AC 1158.
[113] Note 109, [36].
[114] Which was part of the reason for rejecting the MOPI class-action claim brought on behalf of patients in [2024] EWCA Civ 1516 (note 44).

Likewise, if information has been anonymised it is unlikely to be considered private and engage Article 8.[115]

The concept of 'identity' and identification is thus an essential boundary within data protection, confidentiality and MOPI law. Understanding this concept is key for any secondary use of patients' data conducted on an anonymous basis. Through three publications, this Part explores the line between personal and non-personal data. Due to the 'publication format' of this thesis, two of these publications are in the form of journal articles, while the third is a book chapter.

1.3 Identity as Interference

The approach taken in this thesis ties the concept of identification to the potential for interference with Article 8(1) ECHR— that is, to the potential for information to affect a person's dignity or autonomy, as protected by the right to private and family life. This conception of interference can thus serve as a benchmark to say whether the scope of information laws have been drawn too widely, or narrowly, to serve their fundamental purpose.

It is helpful to ground the conception of identification in the potential for a particular use of data to have impact on a natural person. By making this connection, I am tying the scope of these laws to their ultimate purpose, as manifestations of the underlying right to private life under Article 8 ECHR. If the scope of these laws is drawn so broadly that they become the 'law of everything',[116] this does not focus decision-makers' attention on the processing which poses the highest risk to patients. At the same time, as explored more in the **DeepMind Case Study** in Part 2 of this thesis, drawing the lines too narrowly can open the door to disproportionate or discriminatory uses of patients' data. This emphasis on impact can support a more contextual understanding of identity and identification.

A simple way of formulating the result would be: if people are impacted by the processing of data (because the data allow them to be in some way judged or scrutinised), this should be considered an identification. Conversely, if most people would not feel 'seen' or (mis)represented by a set of information, and it does not permit any decisions to be made about them, there is no privacy-based reason to call that information an identification. This assessment is difficult to make in the abstract, because (per my argument) the capacity for data to impact people in this way depends not only on the nature of the data, but also on the context in which it is used.

This clarification represents a contribution in both in terms of legal doctrine, and in a policy sense. The bare legislative, or caselaw, tests require some gloss to be meaningful. From a policy

[115] As was the view of the UK Government, expressed in Joint Committee on Human Rights, *Legislative Scrutiny: Offender Management Bill (third report)* (2006-7), Appendix 6) (Q3), available from: https://publications.parliament.uk/pa/jt200607/jtselect/jtrights/39/3913.htm.

[116] N. Purtova, 'The law of everything. Broad concept of personal data and future of EU data protection law' (2018) 10 Law, Innovation and Technology 1, 40-81.

point of view, there is also a need to flesh out the justification for treating some information as regulated personal data, and other information as just 'data.' Both morally and intellectually, I am drawing from Article 8 ECHR in my exploration of this distinction.

The purpose of Article 8 ECHR is to prevent unjustified interference with private life. This Part therefore argues that the initial trigger for the application of these laws should be the potential for a particular use of information to interfere with private life. My emphasis on interference leads to a more contextual understanding of identity, and identification, as the potential for impact will vary with the circumstances and purpose of data use. As the 1.pseudonymisation article establishes, a researcher accessing data through a secure platform is much less likely (per my argument) to use information in a personally intrusive way. This is not only due to the technical safeguards of the platform, but also to the strict contractual conditions and (perhaps most importantly) professional norms warding against idle curiosity about individuals, or using data for any reason other than answering a legitimate research query by interrogating trends across a dataset. This is a core aspect of the **Scientific Research Case Study**. This is different from an advertiser seeking to profile individuals within a dataset, as explored in the 3.profiling chapter. Even with information as sensitive as health data, therefore, it is important to note *who* is using a dataset—and with *what agenda*— when evaluating the potential for the processing to identify an individual.

This contextual approach to identification is developed in the first two publications in this Part 1. I will first outline the background of each publication, for the sake of context, and to explain my role in writing each paper. I then explain the contribution each publication makes to my overall thesis.

1.4 Publication 1: Pseudonymisation Article

Publication Title: Are 'Pseudonymised' Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK

Publication Type: Journal Article

Authorship: Miranda Mourby,[117] Elaine Mackey, Mark Elliot, Heather Gowans, Susan E. Wallace, Jessica Bell, Hannah Smith, Stergios Aidinlis and Jane Kaye

Background

This article (which I will call 'the 1.pseudonymisation article') was published in the *Computer Law and Security Review* in 2018, and was co-authored with project colleagues. At the time of writing the article, we all worked on the ESRC-funded Administrative Data Research Network

[117] I am listed as the corresponding author for every co-authored journal article contained in this thesis. The author order given for each publication reflects the order given in the published version.

('ADRN'). As part of this interdisciplinary collaboration, we were encouraged to co-author publications to reflect our work across the project.

Following some discussion with my co-authors[118] on the GDPR's definition of pseudonymisation, I developed the legal arguments, with particular reference to *Breyer v Germany* (which was my idea). I wrote all the text, apart from that contained within the text box in section 2.2 (internal page 5) of the article. The text box outlines a scenario about public authority A, research centre B and researcher C. These self-contained paragraphs were written by Elaine Mackey (a data scientist) who contributed a statistical understanding of identification risk. All co-authors read and commented on, drafts of the article (before submission, and as revised in response to peer review). The ideas and language within the article are otherwise my own, albeit inevitably shaped by discussions with my co-authors.

The title of the article refers to 'administrative data.' This phrase comes from the name of our ADRN project, where it was defined as the information public authorities collect in the course of delivering public services (as opposed, for example, to the employee data they process for their internal functioning). The remit of the article is thus broader than that of my thesis, because it mirrors the scope of the ADRN project. However, the legal question at its core— what is the correct definition of personal data for the purpose of the GDPR?— is still an important starting point for this Part of my thesis. Administrative data is a broad enough concept to encompass 'patient data' as defined in this thesis (data the NHS holds about its patients). Thus, the arguments made in this article still apply to the secondary uses of patients' data.

Contribution of the pseudonymisation article

The core contribution of the 1.pseudonymisation article for this thesis is that it provides a basis to argue that patients' data can be processed anonymously when used for secondary purposes, such as scientific research. This is a key finding for how such secondary uses are regulated, particularly under data protection law.

The 1.pseudonymisation article was written in response to concerns that the GDPR would expand the definition of personal data, and that this would make it difficult (or impossible) for public authorities to re-use data for research on an anonymous basis. This would have had serious policy implications, as much research on public sector data is conducted on anonymous basis. Even outside the NHS, the government was passing (at the time) the Digital Economy Act 2017, of which Part V was dedicated to anonymous use of public authority data for research purposes. As well as clarifying legal doctrine, therefore, there were serious policy reasons to clarify the concept of pseudonymisation in the GDPR.

This article defends the continuing possibility of using public-sector data for research on an anonymous basis, even following the implementation of the GDPR. The argument focuses

[118] Particularly Mark Elliot and Elaine Mackey, who contributed a data science perspective on this concept.

closely on the text of recital 26 GDPR, as well as the definition of pseudonymisation in Article 4(5) GDPR. We argued that recital 26 GDPR contains the Regulation's most important definition of personal data, and that the GDPR's new definition of 'pseudonymisation' in Article 4(5) does not expand the definition of 'personal data' from that which existed in the previous Data Protection Directive.

In sum, and consistent with the arguments presented in this thesis, we argued that personal data are not defined by types of information, such as 'pseudonymised' data. Rather, the concept depends on *the capacity of information to identify people depends on a combination of information and context*. The 1.pseudonymisation article thus adopts an interpretation of the GDPR that is more centred on risk to people than it is on a scientific typology of data. This is important for this Part 1 of my thesis, because these risks of harm align with the potential for interference with private life which (I suggest) should be the benchmark for identification. Rather than equating types of data with risks of harm, the 1.pseudonymisation article focuses on the human context in which it is used, to emphasise the importance of information governance in ensuring that data subjects are not impacted by the use of their 'anonymised' data (a point further developed in the 2.anonymity article, below).

Due to the time the 1.pseudonymisation article was written, we relied on the judgment of the EU Court of Justice ('ECJ') in *Breyer v Bundesrepublik Deutschland.*[119] This is because we were considering whether information could be personal data in the hands of one data controller, but still anonymous when shared with a third party under controlled conditions. The *Breyer* judgment provided, at the time,[120] the best EU-level guidance as to whether information which is personal data for one party should be seen as personal data for all other parties. In other words, whether the quality of anonymity (or identification) is an intrinsic quality of the data, or partly a function of context. We argued that the quality of anonymity (or identification) is partly a function of human context; a position that differed from the purely data-centric.

We found authority for our contextual interpretation of 'personal data' from the *Breye*r judgment, in which the ECJ had to determine whether dynamic IP addresses collected by a media services provider constituted personal data. Rather than making a decision based solely on the information in question, the Court considered whether the provider had 'means reasonably likely to be used' (per recital 26) to link the IP addresses to individuals, including whether they had a legal channel to perform this linkage.[121] Although the *Breyer* case was decided under the Data Protection Directive, it provided an important gloss on the correct interpretation of recital 26 of the Directive, which contained substantively similar phrasing as recital 26 GDPR. This decision had not been reflected in any EU-level guidance on anonymisation, and the Article 29 WP guidance had not been updated. We argued that it *should*

[119] (Case C-582/14) *Breyer v Germany* | [2017] 2 C.M.L.R. 3.
[120] Since then, additional guidance has been provided in *SRB v EDPS* (below, note 122) as well as the European Data Protection Board Guidelines on Pseudonymisation (note 104).
[121] Note 119, [48].

be considered, *before* assumptions were made about the scope of the GDPR and its implications for the re-use of data on an anonymous basis.

Our point has, however, been addressed more directly in the subsequent decision in *SRB v EDPS.*[122] In this 2023 judgment, the General Court overruled the European Data Protection Supervisor's argument that information shared with Deloitte must have been personal data, because they were identifiable for the sharing public authority. The Court followed the previous authority in *Breyer* and ruled that it was necessary to 'put oneself in Deloitte's position' when determining whether the information constituted personal data, as processed by the receiving party.[123] In doing so, they affirmed the application of *Breyer* to pseudonymised data, and largely confirmed the arguments outlined in the 1.pseudonymisation article. The EDPB has since largely followed suit in their pseudonymisation guidelines, by distinguishing between the pseudonymisation process and the conditions for anonymity.[124]

For the sake of this thesis, therefore, the key contribution of this article is its clarification that patients' data *can* be anonymised when used for secondary purposes (such as within the **Scientific Research Case Study**). The next publication explains why I suggest these data *should*, wherever possible, be anonymised when used for purposes other than a patient's care.

1.5 Publication 2: Anonymity Article

Publication Title: Anonymity in EU Health Law: *not* an Alternative to Information Governance

Publication Type: Journal Article

Authorship: Sole

This sole-authored article (the '2.anonymity article') was published in the *Medical Law Review* in 2020. The legal reasoning of the 1.pseudonymisation article has been largely vindicated by the EU General Court in *SRB v EDPS.*[125] But in the 2.anonymity article I sought to address the ethical dimension of the criticisms of anonymisation. The defence ultimately comes down to my conception of privacy as founded in interference with Article 8 ECHR, rather than privacy as control.

Contribution of the Anonymity Article

Following on from the contribution of the 1.pseudonymisation article—that patients' data *can* be used anonymously for secondary purposes—the 2.anonymity article establishes why they

[122] Case T- 557/20 *Single Resolution Board v European Data Protection Supervisor* [2024] 1 C.M.L.R. 46.
[123] Ibid, [97].
[124] Note 104.
[125] Note 122.

should be used on an anonymous basis. It does so by defending the legal concept of anonymity, in UK and EU law.

Anonymisation has been criticised by several authors.[126] These critiques merit thorough reflection, because they touch on more than one of my thesis case studies. The example of a university researcher receiving data through a secure platform on an anonymised basis (per the **Scientific Research Case Study**) is relatively low-risk in terms of capacity or motivation to scrutinise any given individual, hence why I argue that reasonable risk of interference with private life can be eliminated. This argument is also made in the context of the **Tissue Donation Case Study.**

The criticisms of anonymity have more pertinence within higher-risk scenarios, where anonymity could be used to avoid regulatory oversight. For example, if NHS had, in the **DeepMind Case Study**, used anonymisation to 'sidestep the GDPR'[127] when it disclosed patients' information to Google. On the facts of this particular Case Study, this did not occur, as the disclosed data of 1.6 million patients were disclosed on an identifiable basis.[128] But this did not negate the broader validity of the objection that the category of legal anonymity could be (ab)used to take processing out of the scope of data protection law.

In the 2.anonymity article, I acknowledge these concerns as legitimate, but do not agree that eliminating anonymity as a legal category is the answer. I agree with El Emam and others,[129] that removing the category of anonymous data removes the regulatory incentive to protect people's informational privacy to the greatest possible extent. In the **DeepMind Case Study**, if patients' data had been used anonymously this would (in my view) have been preferable for the patients affected by the disclosure. As I will explore in Part 2 of this thesis, Google DeepMind received far more patient data than was necessary for the stated purpose of building an app to detect acute kidney injury.[130] Had the category of anonymous research been used in this case study, the harm of disproportionate and irrelevant disclosure[131] could have been avoided, as all individually disclosive information should have been withheld from disclosure.

When I say that anonymity protects people's privacy, I therefore mean in the sense of protecting their identity. This is because 'identification' is a central part of data protection law, and my starting point for defining privacy. There are other ways to define privacy that do not derive from the ECHR. In particular, various scholars have referred to a conception of privacy as the

---

[126] For example, Andrew and Baker (note 135, below), and Johnson et al's (note 127 below) suggestion that anonymisation allows people to 'sidestep the GDPR' and, perhaps most influentially, P. Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701.
[127] As suggested as a hypothetical (but not actual) possibility by T. Johnson, K. Kollnig and P. Dewitte, 'Towards responsible, lawful and ethical data processing: patient data in the UK' (2022) 11 Internet Policy Review 1, 1-25.
[128] See Hodson, note 47.
[129] K. El Emam and C. Álvarez, 'A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques' (2015) 5 International Data Privacy Law 1, 73-87.
[130] See Hodson, note 47.
[131] E.g. of information about abortions and HIV status, see Part 2 of this thesis.

ability to control one's personal data,[132] but the English courts have rejected loss of control over one's private information as a harm protected by data protection law.[133] Although autonomy and control *do* appear to be protected under the emerging MOPI tort,[134] as recognised among the values protected by Article 8 ECHR, these values are not at the core of the conception of privacy I use in this thesis. I am centring the idea of 'interference' and impact in my legal doctrinal conception of privacy, as the key touchstone for the engagement of Article 8. Ultimately, this is because 'interference' comes from the text of Article 8 itself, which prohibits any 'interference by a public authority with the exercise of this right except such as is in accordance with the law.'

By defending anonymity on the basis that it prevents interference with privacy rights, I am therefore using a conception of privacy that is founded in interference with private life, rather than individual autonomy. A 'privacy as control' model of privacy, which prioritises informational self-management above other values within Article 8 ECHR, could hold that personal data should never be anonymised, because then the data subject always retains the theoretical possibility to control their information via their data subject rights.[135] On this conception of privacy, anonymity would actually constitute a *loss* of privacy, because it takes away the possibility for control.[136] As discussed in the Introduction to this thesis[137] (and again in the 6.EHDS chapter[138]), this control is not worth prioritising because it is, in fact, often theoretical, with the dispersed volume of personal data exceeding what any of us can effectively control. Neither is this model equitable in practice, as even the limited control it *can* offer is only realistically available to time-rich, data-educated people with the resources and motivation to pursue their informational self-determination.[139] Finally, on a communitarian understanding of privacy,[140] there is broader social good in preserving a category of information which can be used for purposes such as medical research, without impacting on individuals.

---

[132] For example, P. Schwartz, 'Internet Privacy and the State' (2000) 32 Connecticut Law Review, 815-859; A. Allen, 'Privacy-as-Data Control: Conceptual, Practical and Moral Limits of the Paradigm' (2000) 32 Connecticut Law Review, 861-875; M. Birnhack, 'In Defence of Privacy-as-Control (Properly Understood) (2025) 65 Jurimetrics (forthcoming).

[133] Particularly in the judgment of Simon Brown LJ in *R. v Department of Health Ex p. Source Informatics Ltd (No.1)* [2001] Q.B. 424 (1999) at [34], but also *Lloyd v Google LLC* [2021] UKSC 50; [2022] A.C. 1217 at [109].

[134] Ibid [2021] UKSC 202, [104].

[135] For example, J. Andrew and M. Baker argue that 'anonymization and pseudonymization raise significant concerns over the ownership of behavioral (sic) data' in 'The General Data Protection Regulation in the Age of Surveillance Capitalism' (2019) 168 Journal of Business, 565–578.

[136] K. O'Hara has also argued that 'privacy as control' is (in this sense) paradoxical, because if one used one's control over information to broadcast it indiscriminately, then one would have control, but very little privacy as commonly understood—see *The seven veils of privacy: how our debates about privacy conceal its nature* (2023, Manchester*:* Manchester University Press).

[137] Introduction, Section VI, subsection C.

[138] Publication 6 **(thesis page 76)** internal page numbers 8-9 and 18.

[139] A point emphasised by S. Delacroix and N. Lawrence (Introduction, note 77).

[140] See A. Allen, note 132.

In asking whether individual patients have been identified, for the purposes of the GDPR and for Article 8 ECHR, I am thus de-emphasising individual autonomy, and focusing on the potential for a use of people's information to affect an individual's dignity, or freedom from discrimination.[141] This point is developed further in publication 3.

1.6 Publication 3: Profiling Chapter

Publication Title: Identity, Profiles and Pseudonyms in the Digital Environment

Publication Type: Book Chapter

Authorship: Miranda Mourby and Elaine Mackey

Background

This book chapter, which I will call the '3.profiling chapter,' was published in Bart van der Sloot and Sascha van Schendel's edited collection *The Boundaries of Data* in 2024.

The idea for the 3.profiling chapter grew out of a single paragraph towards the conclusion of the 1.pseudonymisation article. The latter focused primarily on 'identifiable' personal data—that is, information which does not directly identify people on its own, but can identify them when linked to their 'real-world' identifiers through means reasonably likely to be used. But, as we noted towards the end of the 1.pseudonymisation article, linkage of information to a 'real-world' self is not the only way of identifying people. Some information is sufficiently intrusive in the hands of a public authority to constitute an identification in and of itself, without the need for further linkage. We acknowledge this in the following paragraph:

> It appears from guidance issued by the ICO in relation to the Data Protection Act 1998 that if a public authority tracks an otherwise unnamed and unknown individual (for example, through CCTV cameras) this is sufficient identification for data protection principles to apply. Similarly, it has been argued that the use of online data to profile or target individuals should constitute use of personal data, even if the individuals in question are unnamed. It may therefore be necessary, if administrative data are to be considered anonymous when used by a researcher, to ensure through safeguards that the researcher does not profile any individual within the data.[142]

One of the anonymous reviewers noticed this point, and highlighted that it could be developed further. I agreed, but felt that information which is intrinsically identifying, without the need for a 'means reasonably likely to be used' to identify people, was its own subject, which warranted a separate piece of work. I therefore explored the question of 'identified' personal

[141] Which are central principles of the ECHR as a whole, see C. Heri (note 2).
[142] Publication 1 **(thesis page 55)** internal page number 11.

data— data which are intrinsically identifying, without the likelihood of further linkage— in this '3.profiling chapter.'

I drafted the chapter, with Elaine Mackey reviewing and commenting on multiple drafts. She provided her expertise as a data scientist and (by then) information governance manager. I provided the legal analysis, and arguments on the caselaw. As editors of the collection, Bart van der Sloot and Sascha van Schendel also reviewed a draft of the chapter, and made comments, prior to accepting it for publication.

By 2022, when I began writing this chapter, four years had passed since the publication of the 1.pseudonymisation article, and we had the benefit of the Belgian data protection authority's decision in the IAB Europe case, as well as the *SRB v EDPS* judgment.[143] The advertising sector falls outside this thesis's scope of secondary uses of health data. Nevertheless, it helps to illustrate a broader point, which is very relevant: that profiling is innately a form of identification, because it interferes with an individual's privacy without the need for any further linkage to them.

Contribution of the Profiling Chapter

The core contribution of the 3.profiling chapter, for this thesis, is to establish that patients should only be considered 'anonymous' within their data if they cannot be individually judged, evaluated or profiled by the information.

Having defended the distinction between anonymity and identity in the 2.anonymity article, the 3.profiling chapter explores the concept of 'identity' in more depth. Per the overarching 'privacy-based' approach of this thesis, it explores the concept by aligning it with the capacity for interference with Article 8 ECHR. This publication uses profiling as an example of processing which is sufficiently intrusive to constitute an interference with the right to private life.

By considering profiling as a form of identification, the 3.profiling chapter thus sheds further light on the working definition of 'identity' in this thesis. Where the information in question is sufficiently unique to that individual, and permits the construction of a profile about them, it does not need to be attributed back to its source to constitute an identification.

In drawing a distinction between pseudonymised and profiling data, this book chapter reveals a central difference between the two. Pseudonymised data, following the decision in *SRB v EDPS*,[144] do not permit individuals to be uniquely scrutinised within a dataset, and judgments made about them. Profiling data, however, put the individual data subject in a position where they could be a commercially valuable proposition, with their profile saleable to potential online advertisers (for example). As this publication discusses, this is the beginning of an

[143] Note 122.
[144] Ibid.

interference with an individual's intrinsic worth (i.e., their dignity[145]) and opens them up to discriminatory judgments. As such, it is consistent with the privacy-based approach of my thesis to frame profiling as a form of identification, as it is an inherent form of interference with an individual's private life.

1.7 Overall Contribution of Part 1

The research question of this thesis is 'can a privacy-based approach help regulate secondary uses of patients' data?' Part 1 answers the question in the affirmative, at least in terms of how a privacy-based approach can help clarify and deepen our understanding of what 'identity' means in the context of patients' data.

Identification is a key initial question in determining the scope of personal data. As well is having a clear importance from a legal doctrinal perspective, its meaning is also significant from a policy perspective. Most research (a key secondary use of patients' data) in the NHS is conducted on an anonymous basis. The third-party anonymisation model, defended in the 1.pseudonymisation article and the 2.anonymity article, is still the basis on which NHS England shares most patient data for research purposes.[146] The 3.profiling chapter then adds the nuance that, if the NHS uses patients' information 'anonymously,' this data should not be so individually detailed that it permits profiling.

1.8 Conclusion on Part 1

In conclusion, a privacy-based approach to identity can help support a more coherent legal conception of what 'identity' means within identification. This can provide a basis for understanding when patient information that is sufficiently intrusive that it cannot be called anonymous (i.e. profiling data) and information which can be considered anonymous in certain contexts (i.e., pseudonymised data).

This is important for the regulation of secondary uses of patients' data, particularly when it is used for research. In particular, it helps to delineate between pseudonymised health data, which do not permit individual profiling, and information which is sufficiently granular about individuals that they can be evaluated according to their particular combination of factors. This supports the idea of a spectrum of identification,[147] which could be simplified as:

---

[145] See Heri, Introduction note 2.
[146] See Introduction, note 34.
[147] Identification is often represented as a 'spectrum,' see for example Understanding Patient Data, 'Identifiability Demystified' (5 April 2017) https://understandingpatientdata.org.uk/sites/default/files/2017-07/Identifiability%20briefing%205%20April.pdf.

1) Attended hospital 4 times

2) Attended hospital 4 times for a respiratory condition

3) Attended hospital 4 times for a respiratory condition, linked to a genetic disorder and a change of accommodation

4) Attended hospital 4 times for a respiratory condition, linked to a genetic disorder and a change of accommodation (London Borough Tower Hamlets). At moderate risk of domestic violence

The first example is clearly not a unique set of information: it is such a common variable that, by itself, it cannot represent any one individual sufficiently to interfere with their privacy. The same can probably said of the second example. By the third and fourth example, however, the confluence of variables begins to build a picture of this individual. Whether accurate or not, there is a representation of this individual that could invite speculation about whether these details listed about them bear any relationship with each other, and what this person might be like. At this point, the individual is no longer of interest purely as one datapoint among many—in which case, they are pseudonymised, and only of informational value as a small part of a pattern. The line between anonymity and identity is crossed when the individual can be scrutinised, evaluated and judged in their own right, as this is in and of itself an interference with Article 8 ECHR.

It is important to clarify that by defending the 'anonymous' use of patients' data, I do not mean that no justification is required when anonymous data are used for secondary purposes. Publications 1 and 2 both emphasise that anonymous data can only be created by processing personal data. When personal data are processed to be used in anonymous research, the justification of this research, and its potential downstream consequences, should be considered at this 'upstream' juncture. For example, in Part 3 I will outline a 'DPIA+' model to evaluate the data protection and equality law considerations of using patients' data for secondary purposes. This could be completed prior to the anonymisation of patients' data, for a programme of secondary uses, and subsequent data access requests could be reviewed considering the guidelines set within this initial assessment (e.g., as to whether there are any limits to the types of uses for the anonymous data). As such, my approach to identification in Part 1 is consistent with the emphasis on evaluation and justification in Parts 2 and 3 of this thesis.

Overall, this thesis considers secondary uses of data using a 'privacy-based' approach. Part 1 explores identification as a vital first dimension of interference with private life. I conclude that risk of identification is contextual, and dependent on the circumstances of processing. I

then explain that this does not create a regulatory loophole, through which the use of anonymisation ‘techniques’ can negate the need to evaluate, mitigate and justify the impact of processing. Rather, the contextual nature of the risk has the opposite effect. If the risk of identification is contingent on who is using the data, for what purposes, for how long, and in combination with what other knowledge or information, this is all the more reason for careful evaluation.

The net contribution of Part 1 is thus a clarified conception of ‘identity’ for the purposes of English information law, which provides a more coherent basis for determining when patients’ data can (and should) be used anonymously. This clarified account of identification can, indeed, help regulate secondary uses of patients’ data.

Publication 1

Publication Title: Are ‘Pseudonymised’ Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK

Short Title: ‘Pseudonymisation Article’

Page Number of Thesis: 55

The following pages incorporate the published PDF of this journal article, as it has been made available online on an open-access basis. The formatting, pagination and footnote numbering of the publication have been retained, and thus the numbering is self-contained.

The pagination sequence of this thesis will then resume at the end of the publication.

Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK

Miranda Mourby [a,*], *Elaine Mackey* [b], *Mark Elliot* [b], *Heather Gowans* [a], *Susan E. Wallace* [a,c], *Jessica Bell* [a], *Hannah Smith* [a], *Stergios Aidinlis* [a], *Jane Kaye* [a]

[a] *Centre for Health, Law and Emerging Technologies ('HeLEX'), Nuffield Department of Population Health, University of Oxford, UK*

[b] *School of Social Sciences, University of Manchester, UK*

[c] *Department for Health Sciences, University of Leicester, Leicester, UK*

Keywords:
Data protection
Privacy
Anonymisation
Pseudonymisation
Administrative data
Re-identification
General Data Protection Regulation

A B S T R A C T

There has naturally been a good deal of discussion of the forthcoming General Data Protection Regulation. One issue of interest to all data controllers, and of particular concern for researchers, is whether the GDPR expands the scope of personal data through the introduction of the term 'pseudonymisation' in Article 4(5). If all data which have been 'pseudonymised' in the conventional sense of the word (e.g. key-coded) are to be treated as personal data, this would have serious implications for research. Administrative data research, which is carried out on data routinely collected and held by public authorities, would be particularly affected as the sharing of de-identified data could constitute the unconsented disclosure of identifiable information.

Instead, however, we argue that the definition of pseudonymisation in Article 4(5) GDPR will not expand the category of personal data, and that there is no intention that it should do so. The definition of pseudonymisation under the GDPR is not intended to determine whether data are personal data; indeed it is clear that all data falling within this definition are personal data. Rather, it is Recital 26 and its requirement of a 'means reasonably likely to be used' which remains the relevant test as to whether data are personal. This leaves open the possibility that data which have been 'pseudonymised' in the conventional sense of key-coding can still be rendered anonymous. There may also be circumstances in which data which have undergone pseudonymisation within one organisation could be anonymous for a third party. We explain how, with reference to the data environment factors as set out in the UK Anonymisation Network's *Anonymisation Decision-Making Framework*.

* *Corresponding author.* Centre for Health, Law and Emerging Technologies ('HeLEX'), Nuffield Department of Population Health, University of Oxford, UK.

E-mail address: miranda.mourby@dph.ox.ac.uk (M. Mourby).

https://doi.org/10.1016/j.clsr.2018.01.002

The forthcoming General Data Protection Regulation ('GDPR')[1] is poised to have wide-ranging impact on those who work with data – how much impact will naturally depend on its interpretation in practice. Whether and in what circumstances de-identified data can be anonymous is an issue of great practical importance for data controllers, but one which has not escaped controversy, particularly given the ambiguity surrounding the concept of pseudonymisation.

Article 4(5) GDPR defines pseudonymisation as the processing of personal data in such a manner that they can no longer be attributed to a specific data subject without the use of additional information, with technical and organisational measures to ensure that they are not attributed to an identified or identifiable natural person. While the GDPR was in its development, some commentators predicted negative implications for research if a subset of 'pseudonymous' personal data was introduced,[2] and even after the final version has been published there appears to be a tendency to regard data as personal if they resemble data which have undergone a process of pseudonymisation.[3]

Instead, however, the GDPR defines pseudonymisation as an act of processing, and not as a category of personal data. It is therefore inadvisable to use the definition of pseudonymisation to determine whether data are personal data. We suggest that the following two-stage reasoning should be followed:

1) Are natural persons identifiable within the meaning of Recital 26, taking into account all the means reasonably likely to be used?
2) If the answer to the above question is yes, has 'pseudonymisation' been applied within the meaning of Article 4(5) GDPR?

The first section of this article explores the concepts of pseudonymisation and anonymisation under the GDPR. We will then examine the importance of anonymisation in potentially sensitive areas such as administrative data research; i.e. research undertaken using data held by public authorities in connection with their functions.[4] Finally, we will consider how anonymisation can be achieved under the GDPR, with reference to the 'data environment' factors set out in the *Anonymisation Decision-Making Framework*.[5] Anonymisation under the GDPR is, we suggest, still possible for key-coded data, and even data which have undergone pseudonymisation per Article 4(5)[6] may be anonymous when shared with a third party.

1. GDPR pseudonymisation and anonymisation

1.1. *Pseudonymisation: GDPR vs 'conventional'*

Article 4(5) GDPR defines pseudonymisation as:

> *the processing* ***of personal data*** *in such a manner that* ***the personal data*** *can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that* ***the personal data*** *are not attributed to an identified or identifiable natural person.*

As the emphasis added above illustrates, the definition evidently envisages that the data in question begin and end the process as personal data. Personal data are defined as data '*relating to*' an identified, or identifiable, data subject.[7] The data processed per Article 4(5) evidently still *relate* to an identifiable natural person; pseudonymisation merely prevents the *attribution* of the data to a natural person. In other words, GDPR pseudonymisation prevents direct identification through attribution, but not through any other means reasonably likely to be used to identify an individual, which must be excluded before he or she is no longer considered to be identifiable.[8]

The word 'pseudonymisation' in the GDPR thus refers to a process which reduces the risk of direct identification, but which does not produce anonymous data. Pseudonymisation is referred to as a means of reducing risks to data subjects,[9] and as an appropriate safeguard for any personal data used for scientific, historical or statistical research.[10] Personal data which have undergone pseudonymisation are within scope of the GDPR, and the data subject rights set out in Articles 15–20 still apply.[11]

[1] Regulation 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/ (General Data Protection Regulation) [2016] OJ L119/1, which will be cited as 'the GDPR'.

[2] Leslie Stevens, 'The Proposed Data Protection Regulation and its Potential Impact on Social Sciences Research in the UK' [2015] EDPL 107.

[3] Matthias Berberich and Malgorzata Steiner 'Blockchain Technology and the GDPR – How to Reconcile Privacy and Distributed Ledgers?' [2016] EDPL 424.

[4] This definition of 'administrative data' is taken from s.64 Digital Economy Act 2017, which provides new powers of disclosure for public interest research.

[5] Mark Elliot, Elaine Mackey, Kieron O'Hara and Caroline Tudor, *The Anonymisation Decision-Making Framework* (UKAN, 2016).

[6] The GDPR does not use the word 'pseudonymous' or 'pseudonymised', although the word 'pseudonymised' has been used by the Article 29 Working Party in their Guidance WP260 on Transparency under the GDPR. For the most part we will refer in this paper to 'data which have undergone a process of pseudonymisation', or similar. If, for ease of expression, the term 'GDPR pseudonymised data' is used in this paper, it is only as a shorthand for 'data which have undergone a process of pseudonymisation'.

[7] GDPR, Article 4(1).

[8] GDPR, Recital 26, as discussed in more detail in section 2.3.

[9] GDPR, Recital 28.

[10] Article 89 & Recital 156.

[11] It is possible, however, that use of pseudonymised data may fall within Article 11 GDPR – processing in which it is not necessary to identify the data subject – in which case these data subject rights may not apply, see Article 29 Working Party *Guidelines of transparency under Regulation 2016/679* WP260, para 57.

The GDPR definition of pseudonymisation differs significantly from the conventional way in which the term has been used. For example, the Anonymisation Decision-Making Framework defines pseudonymisation as:

> *A technique where direct identifiers are replaced with a fictitious name or code that is unique to an individual but does not itself directly identify them.*[12]

Similarly, the Information Commissioner's Office defines pseudonymisation as:

> *The process of distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity.*[13]

These orthodox articulations of pseudonymisation fall short of what is required within pseudonymisation under the GDPR. The GDPR does not merely describe a technique or a process – in fact, it does not specify at all what techniques should be used, other than stating that a 'process' must be applied. GDPR pseudonymisation requires not just a process but an ultimate 'success state', in which the data cannot be attributed to an individual without the use of additional information. Even this additional information is addressed within the definition, as it must be subject to '*technical and organisational measures*' to prevent reattribution. Thus, the data must not only be modified so that they are not directly identifiable, they must also be protected against re-identification.

The significance of this discrepancy is that conventional characterisations of pseudonymisation are neutral as to whether the resulting data are personal or anonymous; whereas under the GDPR, the pseudonymised data are personal, but protected against identification. This raises the question of whether 'pseudonymisation' can still be discussed as part of an anonymisation process, or whether all 'pseudonymised data' must be considered personal once the GDPR is in force.

1.2. *Pseudonymisation as anonymisation?*

In 2014, the EU Article 29 Working Party produced guidance to the effect that:

> *pseudonymisation is not a method of anonymisation. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure.*[14]

This is, in essence, the meaning of 'pseudonymisation' within the GDPR. In combination with the stipulation elsewhere in the 2014 guidance that de-identification must be 'irreversible' to be considered 'anonymisation', it appears to form a strong case against the use of pseudonymisation to create anonymous data, especially where original identifying information is retained.

It is trite to say that data which have undergone pseudonymisation in the GDPR sense are personal data; this is merely a function of the definition given in Article 4(5). However, as noted above, the word 'pseudonymisation' is not always afforded the same meaning. The UK Information Commissioner's Office ('ICO') suggests that pseudonymisation *can* produce anonymised data on an individual-level basis.[15] While it acknowledges that this data may pose a greater privacy risk than aggregate anonymous data, they ultimately conclude:

> *This does not mean though, that effective anonymisation through pseudonymisation becomes impossible.*[16]

The ICO has maintained this position in relation to the GDPR, advising on its website:

> *Personal data that has been pseudonymised – e.g. key-coded – can fall within the scope of the GDPR depending on how difficult it is to attribute the pseudonym to a particular individual.*[17]

This is to say, data which have been pseudonymised *can* fall within the scope of the GDPR, i.e. they *can* be personal data, but this is not necessarily the case. Conversely, if they *can* fall outside the scope of the GDPR, then it must follow that they can be anonymous. 'Pseudonymised' in this context evidently means 'key-coded,' or the equivalent. This guidance may contrast with the definition of pseudonymisation within the GDPR, but makes sense if the word 'pseudonymised' is understood merely as a de-identification technique for individual level data – or the 'conventional sense' as it is referred to in this paper.

The ICO's guidance on the status of pseudonymised data under the GDPR has been adopted by the Government in the Explanatory Notes to the current Data Protection Bill,[18] and will be a key point of reference for UK data controllers in determining whether they are handling personal data. It appears from the ICO's guidance that pseudonymisation in the 'conventional' sense can still be spoken of as creating anonymised data. But what about pseudonymisation in the GDPR sense? Can data which have undergone this type of processing ever be deemed anonymous?

---

[12] Note 5, page 15.

[13] Information Commissioner's Office, *Anonymisation: managing data protection risk code of practice* (Wilmslow, November 2012) < https://ico.org.uk/media/1061/anonymisation-code.pdf> accessed 7 December 2017.

[14] Article 29 Working Party *Opinion 05/2014 on Anonymisation Techniques* WP216 (Brussels, 10 April 2014).

[15] Note 13, page 7.

[16] Ibid, page 21.

[17] ICO, 'Overview of the General Data Protection Regulation (GDPR)', https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/ accessed 6 November 2017.

[18] Explanatory Notes to the Data Protection Bill 2017-19 <https://publications.parliament.uk/pa/bills/lbill/2017-2019/0066/18066en.pdf> accessed 20 September 2017.

Public Authority A provides administrative personal data to a Research Centre, B, to be used for research purposes. The Research Centre wishes to share this data with Researcher C, but is not sure whether they would be disclosing personal data.
Research Centre B processes the personal data, and removes the information which is deemed to be directly identifying. These identifiers are held separately within Research Centre B, with technical and organisational controls to prevent their reattribution to the research data.
Researcher C accesses the research data in a secure lab (Research Centre B). She has completed the Centre's accreditation training so she knows she cannot bring a phone or tablet into the room where she is working on the data, and the computer she works on is not networked. In addition, she signs an agreement with Research Centre B not to attempt to identify any natural person within the data (she is interested solely in the patterns within the data, which might help their project). All her analytical outputs from her work are checked before she is allowed to take them out of the centre.
Researcher C has no relationship with Research Centre B, or with Public Authority A, which would enable her to access any potentially identifying information. She has no means by which she is reasonably likely (or indeed likely) to obtain the information which would identify the data. The information exists, and so identification is not theoretically impossible and the processing is therefore not technically irreversible. However, it is extremely unlikely that the researcher could or would have access to any information which would enable her to identify natural persons within the data.
Is Researcher C accessing personal data?

The example set out in the Box above illustrates the ambiguity of the relationship between pseudonymisation and anonymisation. Following the 2014 Article 29 Working Party guidance, it would be possible to conclude that the researcher is accessing personal data, as the de-identification processing would not be irreversible, just very unlikely to be reversed by the researcher. To use the definition of pseudonymisation under Article 4(5) as a benchmark would also yield the conclusion that these data are personal; that is simply to say they have undergone a process of pseudonymisation and therefore they are personal.

However, it is the argument of this paper that a more nuanced approach should be adopted. Following the logic of the Court of Justice of the European Union in *Breyer v Germany*, the focus should be on the relationship between the parties, and whether these relationships enable the researcher to identify the data. As explained in more detail below, *Breyer* is authority for the proposition that the scope of personal data should be determined by assessing whether there is a means reasonably likely to be used to identify individuals, and not merely a theoretical possibility of identification.

As it is Recital 26 of the GDPR, and not Article 4(5), which determines whether the data are personal data, this leaves open the possibility that data which have undergone GDPR pseudonymisation could be anonymous for a third party such as Researcher C.

In short, 'conventional' pseudonymisation should still be available as an anonymisation technique under the GDPR, however for the reasons outlined in the previous section it may preferably be described by an alternative term, such as 'de-identification', in order to delineate it from the GDPR definition.

As to whether anonymisation is possible for GDPR pseudonymised data,[19] it is necessary to consider the case of *Breyer* in more detail.

1.3. *GDPR anonymisation:* Breyer v Germany

To understand when data can be considered 'rendered anonymous' under the GDPR, it is necessary to consider the detail of Recital 26. In its final form, Recital 26 GDPR reads as follows:

> *The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an 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. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.*

While its length and detail do not necessarily render it immediately accessible, the correct interpretation of Recital 26 is essential to understand the rest of the GDPR. It sets out the test to determine whether a natural person is identifiable, and therefore whether any data are personal and thus within scope of the Regulation. At the beginning of this paper, we suggested that the question of whether data are personal should be separated from that of the application of pseudonymisation; this is to prevent any confusion which might result from the reference to pseudonymisation within Recital 26.

Prior to the finalisation of the GDPR, the ICO warned of the possible confusion stemming from the text of this Recital. Writing on the Council's text of what was then Recital 23, the Commissioner's Office commented as follows:

> *In our view there should be a single definition of 'personal data'. Therefore it is welcome that 'pseudonymous data' is no longer treated as a separate category of personal data. However, pseudonymisation should only be relevant as a privacy enhancing technique – for example in relation to data minimisation or security. It would be better not to try to define pseudonymisation in the context of the definition of personal data.*
>
> *As it stands, the relevant Recital (23) is confusing. It says that pseudonymous data should be considered as information on an*

[19] See note 6 above.

> *identifiable natural person – this implies all pseudonymous data whoever it is held by. However, the relevant Recital's new reference to the likelihood of identification presumably means that some pseudonymous data would be personal data whilst other pseudonymous data would not be, depending on the likelihood of the relevant identifying information being added to the pseudonymous information.*[20]

Unfortunately, despite the ICO's warning, discussion of pseudonymisation *has* been included in Recital 26. This gives the impression that pseudonymisation is determinative of identifiability, rather than a process to be applied to personal data as defined in Article 4(5). This in turn creates confusion between the test to determine whether data are personal, and the definition of whether personal data have been successfully pseudonymised.

For example, the ICO's guidance that key-coded data and IP addresses 'can' fall within the scope of the GDPR contrasts with Matthias Berberich and Malgorzata Steiner's analysis of Blockchain ('BC') data under the GDPR. Citing the then pending case of *Breyer v Germany*,[21] they reason as follows:

> *Whether the use of BC must comply with the GDPR, will first and foremost depend on whether personal data is stored on BC under respective business model. Most of currently discussed use cases involve transactions of all sorts, be it financial assets, property registers or smart contracts, which all usually involve specific information in relation to a specific person as a rightholder, owner or contract party. Albeit this information is normally encrypted and can only be accessed with the correct keys, encryption of the data as such- i.e. giving access only to authorised parties – will normally not take such information out of the scope of the GDPR. Even if personal information only entails reference ID numbers, such identifiers are typically unique to a specific person. While in all such cases additional information may be necessary to attribute information to the data subject, such information would be merely pseudonymised and count as personal information.*[22]

Under this analysis, encrypted data would count as data which have undergone pseudonymisation, because the process of encryption appears to correlate with the process of pseudonymisation under Article 4(5). However, if the two-stage process advocated in this paper were adopted, the relevant questions would be:

1) Can the data held on Blockchain identify natural persons by any means reasonably likely to be used?
2) If so, have they undergone pseudonymisation within the meaning of Article 4(5)?

The answers to these questions will inevitably vary from one situation to the next, but if the data are found to be personal data this should be because they could identify individuals through means 'reasonably likely to be used,' and not because the encryption process resembles pseudonymisation.

The importance of considering such means as are 'reasonably likely' to be used was highlighted by the Court of Justice of the European Union in their ultimate judgment in the *Breyer* case. The case brought by Patrick Breyer against the German Government was referred to the Court of Justice of the European Union ('CJEU') for a determination as to whether the data in question were personal data. The German government held the IP addresses of individuals who had visited public authority websites; while these addresses related to natural persons, they could not be attributed to identifiable individuals without the use of additional information. This additional information was held by Internet Service Providers.

The data in question were not pseudonymised or de-identified; rather, they were partial, and could only be identified by the additional information. In this respect they were analogous to pseudonymised data, although the question posed was whether individuals could be identified through construction, not reconstruction, of a dataset. The debate as to whether the data were personal, therefore, mirrors the debate over whether pseudonymised data are always personal data.

The opinion of the Advocate General essentially corresponds to the argument that pseudonymised data are always personal. It was submitted by AG Sanches-Borodona that data would be personal as long as a known third party held identifying information, which *could* be used to identify the data (regardless of likelihood of attribution). The opinion was criticised by some on the grounds that it represented an absolute approach which would extend the scope of the GDPR too widely, burdening data processing entities in a way which would be incommensurate with the actual risks to the privacy of data subjects.[23]

The opinion of the Advocate General was not, in the event, followed by the CJEU, who favoured what was termed a more 'relative' approach. The CJEU favoured a logic under which the *means* by which this additional data could be used to identify the data were taken into account. In their view, it was necessary to determine '*whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constitutes a means likely reasonably to be used to identify the data subject.*'[24] They concluded that the data were personal, but only because of the existence of legal channels enabling the competent authority to obtain identifying information from the internet service provider in the event of a cyber-attack.[25] In the absence of these channels, the data would not have been considered personal simply because a known third party could identify them:

> *Thus it appears that the online media services provider has the means which may likely reasonably be used in order to identify*

[20] 'ICO analysis of the Council of the European Union text of the General Data Protection Regulation', <https://ico.org.uk/media/1432420/ico-analysis-of-the-council-of-the-european-union-text.pdf> accessed 6 November 2017.

[21] Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI: EU: C: 2016:779.

[22] Matthias Berberich and Malgorzata Steiner 'Blockchain Technology and the GDPR – How to Reconcile Privacy and Distributed Ledgers?' [2016] EDPL 424.

[23] Gerard Spindler and Philipp Schmechel, 'Personal Data and Encryption in the European Data Protection Regulation' [2016] JIPTEC 7 (2).

[24] Note 21 at 45.

[25] Ibid at 47.

the data subject [. . .] a dynamic IP address registered by an online media service provider [. . .] constitutes personal data [. . .] in relation to that provider, where the latter has the legal means which enable it to identify the data subject with additional data which the internet service provider has about that person.[26]

The judgment of the CJEU in *Breyer* is crucial authority on the interpretation of Recital 26 of the Directive, and, by extension, Recital 26 GDPR. While the IP addresses in question were not pseudonymised, but partial, data, the principles on which they were judged to be personal closely mirror those which would apply to pseudonymised data. Unless a drastically different approach is applied to individual level, de-identified data under the GDPR, it should be possible for these data to be personal or anonymous, depending on the circumstances.

While *Breyer* was, of course, a judgment on the existing Directive and not the GDPR, the text of Recital 26 in the GDPR and in the Directive differ very little in substance. The relevant sentence in the Directive reads:

Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person.

The inversion of 'reasonably likely' from 'likely reasonably' in the GDPR does not alter the meaning of the sentence, although it is perhaps more natural English syntax. The only additions to the Recital 26 in the GDPR appear to be the explicit mention of singling out as a method to be considered as a potential means of identification (of which, more below in section 2.3). To introduce a category of data which are always identifiable, irrespective of likelihood of attribution, would represent a significant departure from the Directive, for which one would hope for clearer evidence. The fact that references to 'pseudonymous data', as proposed in 2013,[27] are not present in the final version of the GDPR, suggests the decision was taken not to introduce such a category.

Therefore, if the precedent set by *Breyer* is to be applied to data which have undergone pseudonymisation under the GDPR, it should be possible for these data to be rendered anonymous in some circumstances. To return to the example given in section 2.2 of GDPR pseudonymised data held by a research centre, and then shared with an external researcher ('Researcher C'), these shared data **would not be personal data** for Researcher C. The key point in *Breyer* is whether the relationship between the parties is such that the researcher has any means reasonably likely to be used to identify data subjects. While these data are undeniably personal for the research centre, they will not be personal for the researcher if she has no means reasonably likely to be used to access the identifiers.

In short, even data which have undergone a process of GDPR pseudonymisation may not be personal data when shared with third parties. It is worth noting that this interpretation is consistent not only with the judgment in *Breyer*, but with the view of anonymisation the UK Government has enacted in the Digital Economy Act 2017. Once s.64 of this Act comes into force, the researcher in the example could access this information as long as it is processed prior to disclosure so that:

a) no person's identity is specified in the information, and
b) it is not reasonably likely that any person's identity will be deduced from the information, by itself or together with other information.[28]

This suggests the UK Government is satisfied that 'good enough' anonymisation can be achieved for the disclosure of administrative data, even when the processing is theoretically not irreversible and original identifiers are retained by the data controller. Analysis provided by the Wellcome Trust also supports this position, and coincides with the ICO response to the draft Recital 23 cited at the beginning of this subsection:

Recital 26 can be read that all pseudonymised data should be considered personal data. . . However, the scope of identifiability is qualified by the reference to "means reasonably likely to be used" as under the 1995 Directive. This suggests that there may be cases where pseudonymised data together with a combination of appropriate organisational, legal and technological measures can be considered anonymous data. A proportionate and context-dependent approach would take into account the range of measures used, including pseudonymisation, to determine whether the data is considered to be identifiable. In order to achieve this it is important to consider the text of Recital 26 in full to understand how the scope of the regulation relates to approaches commonly used in research.[29]

Applying this argument in a research context, therefore, even if data are personal and GDPR pseudonymised within a research centre, it is possible for them to be anonymous for a third party researcher.

1.4. *Singling out*

The other respect in which it has been suggested that the GDPR expands the scope of personal data is through reference to 'singling out' in Recital 26. It has been suggested by the authors of the authoritative *Companion to Data Protection Law and Practice* that it is clear that as long as a person can be singled out he or she is regarded as identifiable.[30] As Leslie Stevens argues, if this is the case it would have serious consequences for administrative data research.[31]

Individuals within pseudonymised data can be singled out, in the sense of being distinguished from one another, and as such the question of singling out is relevant for our present

[26] Ibid at 48–49.
[27] Note 20.
[28] Digital Economy Act 2017, s.64(3).
[29] Beth Thompson, 'Analysis: Research and the General Data Protection Regulation,' Wellcome Trust July 2016, page 2 https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf.
[30] Rosemary Jay, *Guide to the General Data Protection Regulation: A Companion to Data Protection Law and Practice* (Sweet & Maxwell 2017) page 339.
[31] Note 2.

purposes. The phrase 'singling out' as used in the GDPR appears to originate from the Article 29 Working Party's 2007 guidance on personal data,[32] although this was built upon previous discussion of 'piecing together' an individual's identity through various pieces of information.[33] Discussion of singling out has led some to suggest that tracking cookies constitute personal data,[34] although this may not be consistent with the CJEU's treatment of IP addresses in *Breyer*.

In her analysis of the proposed drafts of the GDPR, Stevens identifies a possible interpretation under which data in which individuals can be singled out (if not directly identified) are personal data. This interpretation stems from an earlier draft of Recital 26 (or Recital 23 as it then was). The EU Parliamentary draft of Recital 23 suggested that any individual capable of being 'singled out' is identifiable, and his or her data is therefore personal:

> *To determine whether a person is identifiable, account should be taken of all the means reasonably likely to be used either by the controller or by any other person* ***to identify or single out*** *the individual directly or indirectly (emphasis added).*

As Stevens points out, it is clear from this draft that an individual is identifiable as long as he or she is capable of being singled out, as the terms 'to identify' and 'to single out' a person are treated as equivalent. In the final text of the GDPR, however, it appears that singling out is only referred to as a method for identification:

> *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 (emphasis added).*

The difference may be subtle, but the movement of the phrase 'singling out' from the end result ('identification or singling out') to one of the methods to be used to achieve the end result ('singling out to identify') dramatically changes the meaning of the Recital. In the final version of Recital 26, therefore, it is clear (if perhaps tautological) that for a person to be identifiable, it must be reasonably likely they will be identified. Singling out need only be considered if it is a means 'reasonably likely' to be used to achieve this end.

Recital 26 explains that the reasonable likelihood of a particular method being used to identify an individual should be determined in light of all the 'objective factors', including cost of and time required for identification, and the available technology. No one method of identifying an individual is considered 'reasonably likely' to identify individuals in all cases, each set of data must be considered in its own unique set of circumstances. Under this case-specific approach, it appears entirely possibly for de-identified, individual-level data to be rendered anonymous as long as the reasonably likely means of identification have been eliminated. Therefore, the fact that individuals can be singled out (i.e. individually differentiated) within 'pseudonymised data' is not sufficient to render these data personal.

One caveat is worth noting, however. It appears from guidance issued by the ICO[35] in relation to the Data Protection Act 1998 that if a public authority tracks an otherwise unnamed and unknown individual (for example, through CCTV cameras) this is sufficient identification for data protection principles to apply. Similarly, it has been argued that the use of online data to profile or target individuals should constitute use of personal data, even if the individuals in question are unnamed.[36] It may therefore be necessary, if administrative data are to be considered anonymous when used by a researcher, to ensure through safeguards that the researcher does not profile any individual within the data.

1.5. *Summary*

The above subsections have explored how 'conventionally pseudonymised' data can still be 'rendered anonymous' for the purposes of the GDPR, how even data which have undergone GDPR pseudonymisation could be legally anonymous if shared with a third party, and how 'singling out' is a *method* of identification, and not identification itself. These three points are vital to rebutting the suggestion that the GDPR increases the scope of personal data to the detriment of research.

The next section addresses *why* an increase in the scope of personal data would be to the detriment of research, with a particular focus on research conducted using administrative data.

2. Anonymisation of administrative data

Administrative data research is one of many areas where clarity as to whether pseudonymised data are personal or anonymous is essential. 'Administrative data' is a term used to describe the information held by public authorities in connection with their functions. Since the report of the Administrative Data Taskforce in 2012,[37] it has been argued that the UK should do more to exploit the secondary usage of this type of data, which can provide vital evidence of the impact of government policy and help shape future policy-making for the better. This type of research, usually carried out using the pseudonymised microdata of those who use UK public services, must avoid infringing the privacy or confidentiality of

[32] Article 29 Working Party, *Opinion 4/2007 on the concept of personal data* WP136 (Brussels, 20 June 2007).

[33] Council of Europe's T-PD Committee, *Report on the application of Data protection principles of the worldwide telecommunication network*, ((2004) 04 final), point 2.3.1.

[34] Frederik J. Zuiderveen Borgesius, 'Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation' [2016] C.L.S. Rev 256.

[35] Information Commissioner's Office, *Determining what is personal data* (Wilmslow, 2012) <https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf> accessed 3 January 2018, page 8.

[36] Note 34.

[37] *The UK Administrative Data Research Network: Improving Access for Research and Policy*, Report from the Administrative Data Taskforce, December 2012, <www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-administrativedatataskforcereportdecember201_tcm97-43887.pdf> accessed 11 October 2017.

data subjects. Anonymisation is thus a key means of ensuring that such research is conducted legally, ethically, and in such a way that the public would support.[38]

If anonymisation of individual-level administrative data is not possible under the GDPR, the regulatory burdens on research centres will increase, and the publication of findings may be impeded. Crucially, anonymisation could also lose its effectiveness in ensuring there is no breach of common law confidentiality. The distinction between personal and anonymous data has long been used as a benchmark for whether a duty of confidentiality arises,[39] or whether rights under Article 8 European Convention on Human Rights ('ECHR')[40] are engaged.[41] If pseudonymised data can no longer be considered anonymous, the pool of valuable data which remain legally 'safe' to work with could drastically shrink, with commensurate damage to this developing field of public interest research.

2.1. *Confidentiality*

The first and perhaps most important reason why anonymisation is desirable is the duty of confidentiality owed by public authority to the people whose data they hold. This is a duty owed under common law, which is classically formulated as follows: is the information itself of a confidential nature (i.e. is it publicly available?), is it communicated in circumstances importing an obligation of confidence, and has that confidence been breached?[42]

Administrative data is a broad category, encompassing many types of information which may be confidential, and stemming from a range of different relationships which might attract confidentiality of communication (for example, a doctor–patient relationship).[43] Information has been held to be confidential if it relates to family[44] and intimate relationships[45]; some administrative data will inevitably contain such information.

As a general rule, the Government has said it is likely that names and addresses of individuals supplied to public bodies in pursuance of their functions would in some cases amount to confidential information.[46] Individuals can reasonably expect their public sector personal data will be held confidential unless they are given some prior warning to the contrary.[47] Therefore, while there may be exceptions, it is safest to assume that administrative data will attract a duty of confidence. Data are only deemed to be confidential if they contain information about identifiable individuals[48]; identifiability being determined in accordance with data protection law.

Consequently, if data have been anonymised a public authority can justifiably claim that there has been no breach of confidentiality as the data have not been passed onto researchers in an identifiable form. The Court of Appeal's judgment in *R (Source Informatics) v Department of Health*[49] still stands: disclosure of anonymised data to a third party does not constitute a breach of confidence under UK law. Anonymisation thus prevents breach of confidentiality when public sector data are provided to a third party such as a researcher. Conversely, if pseudonymised microdata were inevitably deemed to relate to identifiable individuals, the accusation could be made that sharing these data with researchers constitutes a breach of confidence, even if the researchers had no practical power or motivation to identify the data subjects.

A well-known example of administrative data research gone wrong is *care.data*, where the public backlash was triggered in part by the fear that pseudonymisation alone would not be enough to protect patient identities, particularly as commercial purchasers of NHS data might have the resources to re-identify individual patients.[50] In light of these concerns, GDPR pseudonymisation affords insufficient protection for administrative research data. For example, Latanya Sweeney successfully re-identified the pseudonymised health records of the then Governor of Massachusetts by cross-referencing with publicly available aggregate census data,[51] rather than using his original health records.

By contrast, the broader considerations within GDPR anonymisation would include questions such as: who has access to the data in question, what other information would they have access to, for what purpose are they using the data, and what resources or motive might they have to try to identify data subjects? All of these factors would inform whether individuals can be identified by means reasonably likely to be used, and could equally be of concern to the public in terms of how their data are protected. In this context, therefore, GDPR anonymisation is the more appropriate option.

2.2. *Privacy*

The right to privacy under Article 8 of the ECHR imposes upon the State an obligation to respect the private lives of the individuals with which it deals.[52] This means that any interference with citizens' rights must be proportionate within the meaning of the ECHR. Any public authorities disclosing administrative data for research will therefore need to do so in a way which is compatible with this obligation.

It is possible that, where data have been anonymised, there will be no interference with ECHR privacy rights. It has been

[38] Ibid at 38.
[39] *R (Source Informatics) v Department of Health* [2001] Q.B 424.
[40] Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 1950 Council of Europe European Treaty Series 5.
[41] *South Lanarkshire Council v The Scottish Information Commissioner* [2013] UKSC 55 at 26.
[42] *Coco v AN Clark* [1969] RPC 41.
[43] Sir Roger Toulson and Charles Phipps, *Confidentiality* (3rd ed., Sweet & Maxwell 2012) at 11001.
[44] *Argyll v Argyll* [197] Ch 302.
[45] *Lennon v News Group* [1978].
[46] Department for Constitutional Affairs, *Public Sector Data Sharing: Guidance on the Law* (DCA 2003) page 20.
[47] *W, X, Y & Z v Secretary of State for Health* [2015] EWCA Civ 1034.
[48] See note 49 below.
[49] [2001] Q.B 424.
[50] Lizzie Presser, Maia Hruskova, Helen Rowbottom and Jesse Kancir, 'Care.data and access to UK health records: patient privacy and public trust' (2015) Technology Science <https://techscience.org/a/2015081103/> accessed 11 October 2017.
[51] Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', (2010) 57 UCLA Law Review 1701, 1719.
[52] *Von Hanover v Germany (2005) 40 E.H.H.R.* 1, at 57.

suggested by the Supreme Court in *South Lanarkshire Council v The Scottish Information Commissioner*[53] that if disclosure '*would not enable* [the data recipient] *or anyone else to discover the identity of the data subjects,*' it would be '*quite difficult to see why there is any interference with their right to respect for their private lives.*'[54] This comment may be predicated on the 'firm distinction' between identifiable and non-identifiable data which Aldhouse and others disavow[55] (can one be certain that an unspecified 'anyone else' could not identify an individual's data?). Yet it is, nonetheless, an indication of the usefulness of using anonymised data to protect privacy.

2.3. Criminal offences

Among the benefits listed by the ICO in complying with their Anonymisation Code of Practice is '*minimising the risk of breaking the law and consequent enforcement action by the Information Commissioner's Office (ICO) or other regulators.*'[56] Under s.66 of the Digital Economy Act 2017 it is a criminal offence to disclose any personal information from administrative data received for research purposes. Such activity would already be an offence under s.55 Data Protection Act 1998 if undertaken without the data controller's consent. Under the Data Protection Bill, it will be a criminal offence to re-identify de-identified data without the consent of the data controller.[57] Again, pseudonymisation itself is not enough to minimise the risk of committing offences under these Acts, but governance of who has access to the data, for what purposes, and in what conditions better mitigate against any potential incursion of criminal liability.

2.4. Data protection

Under the current UK data protection law, set out in the Data Protection Act 1998, pseudonymised data are like any other type of data, and are subject to the data protection principles unless they do not identify a living individual, either on their own or in conjunction with other information which is (or is likely to be) in the possession of the data controller.[58] The ICO views pseudonymisation as one technique capable of producing anonymised data,[59] but does acknowledge that pseudonymised data may be particularly vulnerable to identifying individuals through the 'matching' of various pieces of 'anonymised' information.[60] Under the GDPR, data which have undergone pseudonymisation may be exempt from certain data subject rights, such as subject access, correction, erasure and data portability requests, as long as the controllers can demonstrate that they themselves are not in a position to identify the data subject.[61] However, while the data remain personal, the relevant data controller will still have a duty to keep records of processing, to carry out privacy impact assessments, to appoint a Data Protection Officer and to demonstrate compliance with the principle of privacy by design.[62] It is arguable whether these data protection obligations are in practice any less burdensome than the requirements of maintaining anonymity under the GDPR.

2.5. Summary

There would be a number of negative implications for administrative data research if the scope of personal data was increased by the GDPR: potential breaches of confidentiality, greater infringement of privacy and increased regulatory obligations. Additionally, if anonymisation were to be supplanted by pseudonymisation as defined in Article 4(5) GDPR, this would restrict the consideration of identification risk to the narrow set of 'technical and organisational measures' required by the definition. This argument is explored further in the next subsection.

2.6. Anonymisation vs pseudonymisation

To return to the definition of pseudonymisation – it is clear that not only is it a very specific definition, but it also requires a very specific form of data protection:

> *The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.*

While this requires '*technical and organisational measures*', it is the '*additional information*' which must be '*subject*' to these measures. It is thus the additional identifiable information, held separately from the pseudonymised data, which must be protected. Therefore, the only risk of identification mitigated against within GDPR pseudonymisation is the risk of identification through the original data held by the controller (or by a third party). Fuller clarification of this definition of pseudonymised data can be found in Recital 29 GDPR:

> *In order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should, whilst allowing general analysis, be possible within the same controller when that controller has taken technical and organisational measures necessary to ensure, for the processing concerned, that this Regulation is implemented, and that additional information for attributing the personal data to a specific data subject is kept separately (emphasis added).*

Reading the two provisions together, it is clear that the only protection required for pseudonymised data is against what

[53] [2013] UKSC 55.
[54] Ibid at 26.
[55] Francis Aldhouse, 'Anonymisation of personal data: a missed opportunity for the European Commission' [2014] C.L.S.Rev. 403.
[56] Note 13.
[57] Data Protection HL Bill (2017-19) 66, cl 162.
[58] Data Protection Act 1998, s.1(1).
[59] Note 13 at 7.
[60] Ibid at 21.
[61] GDPR, Article 11.
[62] GDPR, Articles 30, 35, 36, 37 and 25.

could be termed the 'internal' risk of identification from additional information retained by the data controller, or by a known third party. It is understandable that only this limited protection is expected for pseudonymised data; unless they are also 'rendered anonymous', pseudonymised data are still treated as identifiable and within the scope of the GDPR[63] and are subject to corresponding (if potentially modified) data subject rights.[64] It is logical, and consistent with existing legislation, to require broader protection before data are considered out of the scope of data protection principles.

The GDPR acknowledges that pseudonymisation is not a complete solution to issues of data protection, clarifying that its explicit introduction in the Regulation is not intended to preclude any other means of data protection.[65] Pseudonymisation alone may be sufficient where it is acceptable for the data to remain identifiable within the meaning of the GDPR; there may even be circumstances in which it is positively desirable to retain the option of re-identifying the data at a later date.[66] However, there will also be circumstances in which data are used on the understanding that the risk of identification will be minimised. The overall risk of identification is not necessarily limited to the risk posed by the original identifying data, but to any means 'reasonably likely' to be used to identify the data, in which case GDPR anonymisation becomes pertinent.

Under GDPR anonymisation, it must be determined whether natural persons are identifiable by any means 'reasonably likely' to be used. These means may include identifying the data through known additional information which is held separately, but may also comprise more indirect methods of identification, such as singling out individuals and determining their identity using other, possibly multiple, sources of information. For example, following the AOL search history disclosure in 2006, searcher no. 4417749 was identified not through the original identifying information held by AOL, but through the content of her searches which revealed her surname, age and geographical area.[67] Whether such detail in pseudonymised data could identify an individual, especially in connection with publicly available information, is a prime example of the additional considerations required as part of the process of GDPR anonymisation.

Anonymisation is thus appropriate in circumstances sufficiently sensitive to warrant broader consideration of identification risk, such as the large-scale use of public sector data for research. Pseudonymisation within the meaning of GDPR Article 4(5) would be inadequate to address all of the risks of identification encompassed within the 'means reasonably likely to be used' test. The breadth and variety of these considerations, and how they can contribute to an assessment of identification risk, are considered in the next section.

3. The data environment

We have, in the previous section, referred to the 'broader considerations' which must be taken into account as part of the process of anonymisation under the GDPR, as opposed to GDPR pseudonymisation. A more detailed and systematic review of these considerations can be found within the Anonymisation Decision-Making Framework ('the ADF'), published by the UK Anonymisation Network.[68]

The ADF is underpinned by the 'data environment' perspective, which shifts the focus from the data alone to the relationship between data and their environment.[69] The data environment is the context for any piece of data, and is considered to be made up of four components:

- Other data: the key question is what (other) data exists in a given data environment? This is what the data controller needs to know in order to assess (some of the conditions for) re-identification risk. Other data consists of co-present databases, data derived from personal knowledge and publically available sources such as public registers, social media profiles etc.
- Agency: there is no re-identification risk without human agency – this may seem like an obvious point but it is one that is often overlooked and the least understood.[70] The key issue here is in understanding who the key protagonists are, and how they might act and interact to increase or negate the risk of re-identification.
- Governance processes: these processes are formally determined in policies and procedures which control how data are managed and, accessed, by whom, how and for what purposes.
- Infrastructure: infrastructure is not dissimilar to governance, it shapes the interaction between data and environment and includes such as operational and management structures as well as hardware and software systems for maintaining security.

Although it is not within the scope of this paper to provide an in-depth analysis of the data environment perspective underpinning the ADF, we will illustrate the way it can help determine the identifiability of data. For this purpose, we apply the four data environment components to the case of the Administrative Data Research Network ('the ADRN').

[63] GDPR, Recital 26.
[64] Note 11.
[65] GDPR, Recital 28.
[66] Luca Bolognini and Camilla Bistolfi, 'Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from Directive 95/46/EC to the new EU General Data Protection Regulation' [2017] C.L.S.Rev. 171.
[67] Michael Barbaro and Tom Zeller Jr., 'A Face Is Exposed for AOL Searcher No.4417749', *The New York Times* (New York, 9th August 2006), as cited by Francis Aldhouse, 'Anonymisation of personal data – A missed opportunity for the European Commission' (note 55).
[68] Note 5.
[69] For more detail see Elaine Mackey and Mark Elliot, 'Understanding the data environment' (2013) 20(1) XRDS 36; Mark Elliot and Elaine Mackey, 'The Social Data Environment' in Kieron O'Hara, M-H. Carolyn Nguyen and Peter Haynes (eds), Digital Enlightenment Yearbook: Social Networks and Social Machines, Surveillance and Empowerment (IOS Press 2014).
[70] Elaine Mackey, 'A Framework for Understanding Statistical Disclosure Processes: A Case Study using the UK's Neighbourhood Statistics' (PhD Thesis, University of Manchester 2009).

The ADRN was designed to be a new legal, safe and efficient pathway for researchers to access linked administrative data for social and economic research. It has secure Research Centres in each of the four countries of the UK. The ADRN provides a useful case study for illustrating how accredited researchers are able to access detailed and sensitive data that are effectively anonymised, or rather what the ADF refers to as *functionally anonymised*.[71] The key point is that the ADRN employs controls on both data and environment to achieve functional anonymisation. Let us consider this in more detail.

Data received by the Research Centre are de-identified (i.e. direct identifiers are removed). Specially trained staff at the Centre prepare the data for the researcher who will access it in a secure room using a non-networked computer. The data that the researcher accesses are considered functionally anonymised. This is because a combination of data and environment controls are enacted to ensure that the re-identification risk is remote. The re-identification risk can be analysed in the context of the four data environment components outlined above: other data, agency, governance processes and infrastructure.

In terms of other data: this may relate to personal knowledge and or external information sources. The ADRN's governance processes and infrastructure around data security, access and use is such that the researcher is unable to bring into the secure environment (other) data. A researcher cannot take any materials in or out of the room in which they access the research data (including mobile phones, memory sticks or pen and paper) and they cannot copy or download data.[72] In addition, all research outputs are checked prior to being released from the Research Centre to ensure that data subjects cannot be re-identified and information about them cannot be disclosed.

In terms of agency: the key agents in this situation are the researchers. The way in which they access the data (as described above) and how they behave in the Research Centre is shaped by the ADRN's governance processes and security infrastructure. A researcher's project proposal is assessed by an independent approvals panel comprising of academics, data providers and members of the public. The panel considers whether the project is of public benefit, scientific merit and feasible. It also reviews the project's Privacy Impact Assessment which is undertaken by the ADRN If the panel approves the project those members of the project's research team who plan to access the data are required to undertake and pass researcher training prior to working in the Research Centre. The training covers: (i) legal issues, (ii) the researcher's responsibilities under the law, (iii) appropriate use of the secure setting and sanctions for misuse and (iv) processes for checking that output are not disclosive.[73] The ADRN controls the who and how of access, ensuring that those accessing the data do so safely, lawfully and responsibly.[74]

Governance processes and infrastructure inform who accesses data, and on what terms. These processes shape users' relationship with data through formal means such as policies and agreements, as well as more informal measures such as de facto norms. The governance provided by the ADRN promotes a culture of responsible use of data, and respect for legal sources of governance such as the Data Protection Act 1998. They also include formal governance through relevant policies, including Privacy Protection and Information Security policy; Safe Users of Research data Environments (SURE) Training Policy; terms of use which specify that the data must be used for research purposes only and the researcher may make no attempt to re-identify any individuals within the data[75]; and a policy which sets out the penalty for any breach of the terms of use.[76]

Although the purpose of these measures is in part to maintain legal anonymity, they would remain pertinent for administrative data research even if a different approach is taken to pseudonymised data. If, contrary to the arguments put forward in this article, it is determined at the UK or European level that all pseudonymised data are the data of identifiable individuals, and cannot be anonymised, good information governance will still be vital to ensure that those data are used legally, ethically and responsibly, and particularly to avoid direct identification of individuals. The importance of information governance is thus not an issue which stands or falls entirely on the GDPR definition of pseudonymised data. However, if anonymisation of pseudonymised data remains a possibility in law, it will help to reinforce good practice by defining identifiability with reference beyond original identifiers, to all the risks which good governance can address.

4. Conclusion

In conclusion, the GDPR should not be seen as expanding or re-defining the scope of personal data through its definition of pseudonymisation. The definition of pseudonymisation is not intended to be used to establish whether data are personal under the GDPR; indeed, it is clear that data to which pseudonymisation is applied are, and remain, personal data. Instead, it is Recital 26 GDPR which must be used to establish whether data are personal. This ultimately poses the question as to whether there exists a means reasonably likely to be used to identify natural persons. As such, anonymisation processes under the GDPR do not necessarily exclude pseudonymisation in the conventional sense, such as key-coding, as long as other environmental controls are in place to prevent the 'data situation' yielding identifiable data.

[71] Functional anonymisation asserts that one cannot determine the status of data as personal or not without reference to their environment, see the ADF at note 5.

[72] For further information on this please see the ADRN website at https://adrn.ac.uk/get-data/secure-access/, accessed 19 September 2017.

[73] <https://adrn.ac.uk/understand-data/sure-training> accessed 19 September 2017.

[74] Ibid.

[75] ADRN Terms of Use <https://adrn.ac.uk/media/174434/adrn021-termsofuse_v00-11_pub.pdf> accessed 19 September 2017.

[76] ADRN Breaches Policy <https://adrn.ac.uk/media/1297/adrn003_breachespolicy_02_pub.pdf> accessed 25 September 2017.

The case of *Breyer* even leaves the door open for data which have undergone pseudonymisation within one organisation, to be rendered anonymous from the perspective of an individual outside that organisation. This judgment demonstrates that the relationship between two parties is key to determining whether identifying information is sufficiently accessible for the data they hold to be deemed personal. This distinction is vital in a research context, as it leaves open the option of sharing 'anonymised' data with researchers, even when data are personal and GDPR pseudonymised within the organisation itself. Close attention to the terms of data sharing agreements will therefore be essential in a research context, in order to share data which would be confidential if identifiable, and to maintain the flow of data for public interest research in future.

Acknowledgements

The authors would like to thank Dr Mark Taylor, and two anonymous reviewers, for their comments on this article.

The authors received support from the ESRC grant number ES/L007452/1 for the Administrative Data Service. The funders played no role in the writing of this article, or the decision to submit it for publication.

Publication 2

Publication Title: Anonymity in Eu Health Law: *Not* an Alternative to Information Governance

Short Title: ‘Anonymity Article’

Page Number of Thesis: 56

The following pages incorporate the published PDF of this journal article, as it has been made available online on an open-access basis. The formatting, pagination and footnote numbering of the publication have been retained, and thus the numbering is self-contained.

The pagination sequence of this thesis will then resume at the end of the publication.

Medical Law Review, Vol. 28, No. 3, pp. 478–501
doi:10.1093/medlaw/fwaa010
Advance Access Publication: May 15, 2020

ANONYMITY IN EU HEALTH LAW: *NOT* AN ALTERNATIVE TO INFORMATION GOVERNANCE

MIRANDA MOURBY*

Centre for Health, Law and Emerging Technologies, Faculty of Law, University of Oxford, Ewert House, Ewert Place, Summertown, Oxford OX2 7DD, UK

ABSTRACT

Data sharing has long been a cornerstone of healthcare and research and is only due to become more important with the rise of Big Data analytics and advanced therapies. Cell therapies, for example, rely not only on donated cells but also essentially on donated information to make them traceable. Despite the associated importance of concepts such as 'donor anonymity', the concept of anonymisation remains contentious. The Article 29 Working Party's 2014 guidance on 'Anonymisation Techniques' has perhaps helped encourage a perception that anonymity is the result of data modification 'techniques', rather than a broader process involving management of information and context. In light of this enduring ambiguity, this article advocates a 'relative' understanding of anonymity and supports this interpretation with reference not only to the General Data Protection Regulation but also to European Union health-related legislation, which also alludes to the concept. Anonymity, I suggest, should be understood not as a 'technique' which removes the need for information governance but rather as a legal standard of reasonable risk-management, which can only be satisfied by effective data protection. As such, anonymity can be not so much an *alternative* to data protection as its mirror, requiring similar safeguards to maintain privacy and confidentiality.

KEYWORDS: Anonymity, Cell donors, Confidentiality, Data sharing, Information sharing, Medical research

* miranda.mourby@law.ox.ac.uk. The author acknowledges support from the Leverhulme Trust project grant RPG-2017-330 'Biomodifying Technologies: Governing converging research in the life sciences,' as well as the Medical Research Council project grant MC_PC_17215 'Developing an informatics supported platform for experimental medicine - MICA, industrial collaboration with Janssen and SomaLogic.' The funders have not had sight of the manuscript and played no role in the decision to submit it for publication. I am grateful to Heather Gowans, Elaine Mackey and Alison Knight for their kind assistance in the development of this paper. I would also like to thank the organisers and attendees of the Medical Research Council's workshop 'Safe Sharing of Research Data: The Role of Legal Agreements When Anonymising' on 25 April 2019. The usual disclaimer applies.

I. INTRODUCTION

Anonymised data are often recognised as an integral resource for biomedical research,[1] but the concept of anonymisation is a source of enduring controversy. Anonymisation has been viewed as an avoidance of regulation,[2] enabling otherwise unlawful or unethical secondary uses of data,[3] and allowing data to be freely used, shared, and sold.[4] The perception that anonymisation is potentially harmful, or fundamentally unworkable, casts doubt on its utility within research,[5] and indeed within areas such as therapeutic cell, tissue, and organ donation which currently require donor anonymity as a default.

In contrast, this article uses the words 'anonymity' and 'anonymisation' to refer to the successful maintenance of anonymity to the relevant legal standard. This standard comes from Recital 26 General Data Protection Regulation[6] (GDPR), which refers to data 'rendered anonymous', meaning individuals are no longer identifiable by any means 'reasonably likely' to be used. It sets a test as to when data can be considered anonymous and out of scope of the GDPR. Meeting this test is better understood as a form of regulatory compliance in itself, as opposed to entry into a Wild West of Big Data flows.

The original contribution of this article is thus to reflect on 'anonymity' with reference to developments within data protection law but also to sources of European Union (EU) healthcare law governing tissue, cells, and organs which also use such terminology but are less studied in connection with anonymisation. These pieces of legislation variously allude to the principle of 'donor anonymity', even though it is obviously not possible for the donor's identity to be unknown to all actors involved in the donation. For this to be feasible, I argue that a more relative interpretation of anonymity is required. The benefit of this relative interpretation is to remove focus away from eliminating identifiability *for the discloser* by modifying the data through an 'anonymisation technique', and instead focus on the capacity of a specific recipient to identify individuals. This draws on existing work on the contextuality, or 'functionality', of anonymisation by Elliot, Mackey, El Emam, and others, identifying legal authority in support of this account, as well as highlighting why the concept is of

1 For example, *R v Department of Health Ex p Source Informatics Ltd (No 1)* [2001] QB 424 (CA) 23.

2 M Bayern, 'DeepMind, NHS Use Anonymized Patient Data in AI to Avoid Regulatory Hurdles' (2018).<https://www.techrepublic.com/article/deepmind-nhs-use-anonymized-patient-data-in-ai-to-avoid-regulatory-hurdles/> accessed 26 February 2020.

3 J Andrew and M Baker, 'The General Data Protection Regulation in the Age of Surveillance Capitalism' (2019) Journal of Business.

4 L Rocher, JM Hendrickx and Y de Montjoye, 'Estimating the Success of Re-identifications in Incomplete Datasets using Generative Models' (2019) 10 Nature Communications 3069.

5 B Clarke, 'Researchers: Anonymized Data Does Little to Protect User Privacy' (2019).<https://thenextweb.com/insider/2019/07/30/anonymized-data-does-little-to-protect-privacy/amp/> accessed 30 July 2019.

6 Regulation 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/ (General Data Protection Regulation) [2016] OJ L119/1, which will be cited as 'the GDPR'

continued relevance, even at a time when high-profile figures suggest that the distinction between personal and non-personal data is obsolete.[7]

By way of background, Section II begins with the Article 29 Working Party (A29WP)'s 2014 guidance on Anonymisation Techniques, and in particular its tension with the realities of biomedical research. Section III then considers the subsequent developments since this guidance was published, such as the transition to the GDPR and the judgment of the Court of Justice of the European Union (CJEU) in *Breyer*.[8] The way in which the scope of personal data was drawn in this case is, I suggest, useful when contemplating 'donor anonymity', which cannot be absolute anonymity. This is explored in Section IV. Section V touches on the forthcoming Clinical Trials Regulation (CTR) and the impact this may have on data sharing.

Finally, Section VI considers the extent to which a clarified understanding of anonymisation addresses criticisms levelled at anonymity within biomedical Big Data. A more relative, organisation-specific account of anonymity could actually help to counter the suggestion that anonymisation is an 'opt-out' from regulation, as it means that data subjects can still exert their GDPR rights against the original provider but can be assured to a reasonable standard that they will not be identified by third parties. However, it is accepted that this does not resolve all potential concerns, and anonymisation may in fact mirror the individually focused weaknesses of data protection law, as well as its strengths.

Ultimately, I argue that anonymity remains a useful and important concept in EU healthcare law, not least because it is unique in indicating a state in which people will not (in all reasonable likelihood) be identified by unauthorised third parties, and this is still a distinction which could matter to patients, donors,[9] and trial participants. However, in order to be of continuing utility it must be thought of in relative terms, and in a way which takes into account information governance as a means of managing context. In healthcare research, where it is possible to regulate behaviour between collaborators through governance measures such as policies and contracts, information of greater research utility can be shared by addressing the lawful behaviour of those given access to the minimised data, and removing reasonable means of identification from them.[10] Biomedical research is, therefore, an area which may particularly benefit from revision of the existing guidance, and a move to a more relative approach.

A. Terminology

For clarity, the term 'relative' is used in this article to reflect the usage by the CJEU to differentiate between different interpretations of the scope of personal data.[11] In

7 F Niker, 'An Interview with Baroness Onora O'Neill (Beyond the Ivory Tower Series)' (6 January 2020) <http://justice-everywhere.org/governance/an-interview-with-baroness-onora-oneill-beyond-the-ivory-tower-series/> accessed 26 February 2020.

8 Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:779.

9 For example, F Mahieu and others, 'Anonymous Sperm Donors' Attitude Towards Donation and the Release of Identifying Information' (2019) 36 Journal of Assisted Reproduction and Genetics 10; G Cohen and others, 'Sperm Donor Anonymity and Compensation: An Experiment with American Sperm Donors' (2016) 3 Journal of Law and the Biosciences 3

10 For more detail on the balance between utility and identification risk, see M Elliot and others, 'The Anonymisation Decision-making Framework' (UKAN, 2016) <https://ukanon.net/ukan-resources/ukan-decision-making-framework/> accessed 6 April 2020.

11 *Breyer* (n 8).

broad terms, the objective approach sees data as personal if 'anyone' could identify them, whereas under the relative approach it is only the actual (be) holder of the information, and anyone they could reasonably approach, who needs to be taken into account.

I have suggested that anonymity is unique as a regulatory standard, as it assures individuals that they will not be identifiable by any means reasonably likely to be used. This is why this article focuses on anonymity, as opposed to 'pseudonymity'. Under the GDPR, pseudonymisation is defined as a process of data minimisation which nevertheless does not prevent data from being personal.[12] This means that, by definition, people within pseudonymised data are still reasonably likely to be identified, and so talking of 'donor pseudonymity' or 'participant pseudonymity' would not have the same meaning. Following the current definition, only if additional steps are taken to ensure that third parties are not reasonably likely to identify individuals can data be said to be anonymous for these people, as opposed to merely pseudonymised.[13] It is a key argument of this article that these additional steps lie in information governance, as opposed to the deletion or aggregation of original data. This is considered in the next section.

II. OPINION 05/2014 ON ANONYMISATION TECHNIQUES

It has already been argued that anonymity in research cannot be objective (or 'idealised' as Saunders and colleagues have characterised it[14]) as there will always be some parties (such as the primary researchers) who will be aware of the participants' identities. Nonetheless, it may well be important for interview participants not to be identified by people outside of this confidential relationship.[15] This section explores how this kind of relative anonymity within research is difficult to achieve in accordance with the A29WP's guidance on Anonymisation Techniques.

The primary legal authority on EU anonymity is Recital 26 GDPR, which provides as follows:

> .... 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. ...

12 GDPR (n 6) art 4(5).

13 M Mourby and others, 'Are "Pseudonymised" Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK' (2018) 34 Computer Law & Security Review 2.

14 B Saunders, J Kitzinger and C Kitzinger, 'Anonymising Interview Data: Challenges and Compromise in Practice' (2015) 15 Qualitative Research 5.

15 ibid.

This establishes a contextual question of fact as to whether data are personal or anonymous, and the GDPR does not specify any set of practices that might be used to remove data from scope.

Nonetheless, there can be a tendency to view anonymisation as a set of practices—rather than a regulatory standard—and when these practices fail, this can lead some to conclude anonymisation itself has failed,[16] rather than that the term has been misapplied to a situation where the reasonable risk of identification has not been excluded, and the test has not been met.[17] It is understandable, however, that the term is used in this way (whether anonymisation has actually been achieved or not) when influential EU-level guidance refers (slightly misleadingly) to 'Anonymisation Techniques'.[18] It is tempting, in the circumstances, to identify anonymisation with the techniques themselves, whether they in fact produce anonymity or not.

The A29WP was a body made up of representatives of supervisory authorities from each EU Member State and adopted its Opinion 05/2014 on Anonymisation Techniques on 10 April 2014. The A29WP has since been replaced by the similarly constituted European Data Protection Board (EDPB), which formally endorsed some of the A29WP's guidance in 2018,[19] but not Opinion 05/2014 itself. While Opinion 05/2014 has not been actively disavowed, and there is no evidence at the time of writing that the EDPB intends to issue any new guidance on anonymisation,[20] its status is less certain than if it had been formally endorsed and is further complicated by intervening developments such as the case of *Breyer* and the transition to the GDPR.

This guidance has still been highly influential (see Subsection II.B), despite the problematic nature of some of its requirements. For the purposes of this article, I will focus on two issues in particular:

1. its neglect of information governance and
2. its suggestion that anonymity requires deletion or aggregation of original data.

A. Information Governance

As regards the first issue, the word 'governance' is used broadly to include any measure which helps to shape the way in which information is used. While the term 'regulation' has equally been used to describe the process of altering others' behaviour by both state and non-state actors,[21] the term 'governance' is preferred here as 'information governance' is a commonly recognised term encompassing the voluntary (and

16 P Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701.

17 Elliot and others also argue that alleged failures of anonymisation are failures of practice, and not of law: M Elliot and others, 'Functional Anonymisation: Personal Data and the Data Environment' (2018) 34 Computer Law & Security Review 2.

18 A29WP, Opinion 05/2014 on Anonymisation Techniques WP216 (Brussels, 10 April 2014).

19 EDPB, Endorsement 1/2018 <https://edpb.europa.eu/sites/edpb/files/files/news/endorsement_of_wp29_documents_en_0.pdf> accessed 27 February 2020.

20 L Taranto and P Garcia, 'Medical Research Council Advises on How to Anonymise Information for Research Purposes' (16 October 2019) <https://www.hldataprotection.com/2019/10/articles/international-eu-privacy/medical-research-council-advises-on-how-to-anonymise-information-for-research-purposes/> accessed 13 November 2019.

21 J Black, 'Regulatory Conversations' (2002) 29 Journal of Law and Society 1.

necessitated) actions taken at organisational level to control how information is used in that particular setting. The word 'governance' has also been used in this broader sense, encompassing leadership, cultures, and relationships, within the context of biobanking.[22]

In contexts such as health and life sciences research, where the relationship between data sharing parties is likely to involve a Data Access Agreement,[23] contracts would be a key example of a governance measure to help achieve, and maintain, anonymisation, by shaping and regulating how data are used. As Stalla-Bourdillon and Knight argue, when a 'dynamic' and environmental approach is taken to anonymisation:

> recipients of anonymized data, although they are not data controllers when they receive the dataset, have to behave responsibly and comply with any licensing obligations imposed by the original data controllers of the raw personal data. Specifically, the former must abide by any licensing limitations upon the purpose and the means of the processing of the data in its disclosed post-anonymization process form to remain outside the scope of data protection laws. At the same time, the characterization of anonymized data should also be dependent upon an ongoing monitoring on the part of the initial data controller of the data environment of the dataset that has undergone anonymization.[24]

An important way to minimise the risk of re-identification of data subjects is thus a contract between data provider and recipient creating legally binding obligations such as a promise not to attempt re-identification, to maintain adequate data security, only to use data for specified purposes, access to be granted only for authorised personnel, records of processing to be kept, and specified procedures to be followed in the case of accidental re-identification. This essentially creates a compromise by providing legal protection for data subjects without exposing them to a reasonable likelihood of identification (which, in effect, data protection law requires for its scope).

In contrast, the A29WP's guidance is implicitly dismissive of the role that contracts can play in enforcing anonymisation. The only mention made comes on page 29:

> State-of-the-art encryption can ensure that data is protected to a higher degree, i.e. it is unintelligible for entities that ignore the decryption key, but it does not necessarily result in anonymisation. For as long as the key or the original data are available (even in the case of a trusted third party, contractually bound to provide secure key escrow service), the possibility to identify a data subject is not eliminated.

22 J Kaye and others, *Governing Biobanks – Understanding the Interplay Between Law and Practice* (1st edn, Hart 2012).

23 See, for example, the standard Data Access Agreement through which information is made available via the European Genome-Phenome Archive<https://www.ebi.ac.uk/ega/submission/data_access_committee> accessed 13 November 2019.

24 S Stalla-Bourdillon and A Knight, 'Anonymous Data v. Personal Data – False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data' (2016) 34 Wisconsin International Law Journal 284.

This does not explore whether, even if the trusted third party was contractually bound not to reveal the key or original data to anyone, or anyone other than an authorised recipient, this key would nonetheless be a means reasonably likely to be used for all other parties. The possibility of identification must apparently be eliminated, not just for the party receiving 'anonymised' information, but for *everyone* for the information to count as anonymised under this guidance. Therefore, even if it would be unlawful, and contrary to contract, for others to obtain assistance in identifying individuals, this is presumably still a 'means reasonably likely to be used' in the eyes of the drafters.

Otherwise, the guidance is largely silent as to the role that due diligence, track-record, contracts, policies, training, line-management, record-keeping, accreditation, Codes of Practice, and audits can play in establishing and maintaining minimal risk of identification, meaning it often reads as 'data-centric' or at least 'provider-centric' rather than accommodating the perspective of the information recipient.[25] The A29WP do advise against 'release and forget' approaches and recommend monitoring and control of risk.[26] However, by focusing the Opinion on techniques such as noise addition, randomisation, k-anonymity, and pseudonymisation, they do not provide any guidance as to how this monitoring should take place, and indeed how this monitoring is itself a safeguard against reasonable means of identification and an important supplement to any data modifying technique. The emphasis, it seems, is not to establish trustworthy governance but to 'eliminate' any technical possibility of identification prior to disclosure. It is, therefore, unsurprising that de-identification of datasets before sharing them has come to be seen as the paradigm for research,[27] rather than anonymisation being understood as an ongoing process of environmental regulation[28] which must be maintained after data have been transferred.[29]

An alternative account of anonymity would see governance of anonymous data as mirroring GDPR compliance in its requirements: accountability, records of processing, purpose limitation, breach procedures, and data minimisation would remain core aspects of risk-management for anonymised data post-disclosure, to ensure no reasonable means of identification re-emerges. It would be helpful if these dimensions of anonymisation were more widely recognised. Controllers entering into a Data Processing Agreement, for example, benefit from helpful guidance, and indeed a checklist within Article 28 GDPR, as to what that contract should cover. The UK Medical Research Council does provide some guidance as to what an Agreement to support anonymisation could include, such as a prohibition on identification attempts and contingency terms in the case of accidental identification.[30] For the sake of data

25 M Elliot and A Dale, 'Scenarios of Attack: The Data Intruder's Perspective on Statistical Disclosure Risk' (1999) 14 Netherlands Official Statistics 6.

26 A29WP (n 18) 24,

27 Rocher, Hendrickx and de Montjoye (n 4).

28 E Mackey and M Elliot, 'Understanding the Data Environment' *XRDS* (2013) <https://dl.acm.org/doi/10.1145/2508973> accessed 16 December 2019.

29 For more detail as to how, see M Elliot and others (n 10).

30 Medical Research Council, Regulatory Support Centre, 'Identifiability, Anonymisation and Pseudonymisation: Guidance Note 5' (2019) <https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-5-identifiability-anonymisation-and-pseudonymisation/> accessed 6 April 2020.

sharing across the EU, this could be developed further at EDPB level and can involve considerations such as:

- Has the discloser exchanged information with the recipient before?
- Is there any reason to doubt their track record of reliable use of data?
- Has the recipient provided evidence of appropriate training or policies as to safe data handling within their organisation?
- Is a data access agreement in place? Does said agreement:
  - prohibit re-identification;
  - prevent all but a specified number of uses;
  - prohibit linkage with other information, or onward sharing of the data;
  - place a time limit on the retention of the data;
  - limit the number of people who can access the data;
  - limit the purposes for which they can use the data;
  - require the recipient to keep records of their use of the data;
  - allow the discloser to audit the use of the data;
  - require the recipient to delete the data at the discloser's request; and
  - make provision for the procedure in the event of accidental re-identification?
- To the extent that preparing the data for disclosure requires internal processing of personal data—has the discloser been transparent with their data subjects about this processing and its ultimate purpose?

The above questions are of the kind discussed at an Anonymisation workshop hosted by the Medical Research Council in 2019, although this is by no means exhaustive.[31] Likewise, while contracts are by no means the sole mechanism through which identifying behaviours could be managed, they provide an example of the way in which information governance can control context, and thus the means reasonably likely to be used by a recipient. It would be valuable if any new guidance from the EDPB could acknowledge this.

B. Deletion/Aggregation of Original Data

The suggestion that original information should be deleted or aggregated is another respect in which the A29WP set an unhelpful precedent, a key quotation from the guidance being:

> Secondly, "the 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

31 'Safe Sharing of Research Data: The Role of Legal Agreements When Anonymising' (*IET*, 25 April 2019). Any errors or misapprehensions are the author's own.

> aggregate the data to a level where the individual events are no longer identifiable, the resulting dataset can be qualified as anonymous.[32]

This passage makes it clear that the A29WP considered data to be personal if they could be identified by 'any' person, including the original controller. This is important as the CJEU implicitly approved, in 2016,[33] the submission that personal data need only be identifiable by those the controller could 'reasonably approach' (see Section III). As such, it is questionable whether this requirement is still tenable. If a data recipient could not reasonably approach the provider of the information for assistance in identifying individuals (eg if it was illegal or contrary to contract), it seems unnecessary and impracticable for the latter to delete their information. This removes from the scope of anonymous data sharing many holders of medical or clinical trial records who naturally cannot delete original data.

Even before the *Breyer* judgment, the deletion requirement was criticised for its impracticality, with El Emam and Alvarez arguing forcefully in 2014:

> The implications of this interpretation are quite severe because some projects and programs will still need the original data to conduct their business. For example, consider a hospital that wished to provide anonymized data for research. The hospital needs to retain the original data because that original data are required to treat the patients. To destroy or aggregate the original data would not make any sense.[34]

This is as true now as it was in 2014, when the A29WP's guidance was first adopted. It has meant that, for the last 5 years, anyone following this guidance would have to inform data subjects that their identifiable information will be shared with third parties, if it is not possible to delete it at source. Yet this guidance remains influential, with the European Medicines Agency (EMA) advising in the context of clinical trial data:

> Pseudonymisation reduces the linkability of a dataset with the original identity of a data subject but when used alone will not result in an anonymous dataset, therefore data protection rules still apply. It is, therefore, important to clarify that pseudonymisation is not an anonymisation method but a useful security measure. Consequently, additional measures should be considered in order to render the dataset anonymised, including removing and generalising attributes or deleting the original data or at least bringing them to a highly aggregated level.[35]

32 A29WP (n 18) 9.

33 *Breyer* (n 8).

34 K El Emam and C Álvarez, 'A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques' (2015) 5 International Data Privacy Law 1.

35 EMA, 'External Guidance on the Implementation of the European Medicines Agency Policy on the Publication of Clinical Data for Medicinal Products for Human Use' (2018) version 1.4, 41 <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-3.pdf> accessed 25 February 2020.

This passage closely echoes the A29WP's guidance cited above in proposing the deletion or aggregation of original data to achieve anonymity, as opposed to controlling the context into which information is disclosed.[36] This proposal is particularly problematic within guidance focused on the *publication* of data, as reducing the prospect of identification from the original data does not affect the ability of others to identify individuals by other means.[37] Also, as the European Commission (EC) has pointed out in relation to serious adverse reactions, it is unrealistic to expect original clinical trial data to be capable of deletion, and so El Emam's objection has thus been tacitly acknowledged by the EC (albeit in the context of deletion where consent has been withdrawn).[38] This highlights another way in which the A29WP's guidance is not the most helpful benchmark for research data. In a controlled access context, where it would be possible to ensure, to a reasonable standard, that the third party cannot or will not identify individuals from the data, it seems unnecessary and undermining of confidence not to be able to tell people that their shared information will be anonymous, especially as there is evidence to suggest that anonymity is an important condition for public approval of health data sharing.[39]

Neglect of information governance and an emphasis on deletion are only two aspects of the A29WP's guidance. However, they are elements which help to explain the perception that anonymisation involves the modification and subsequent neglect of data. As the above citation from the EMA demonstrates, they are also elements which continue to prove influential within EU healthcare practice. While this is problematic in its own right, developments since 2014 also challenge this understanding of anonymity. These developments are explored in the next section.

III. *BREYER* AND THE GDPR

In assessing whether the A29WP's guidance is still a useful benchmark, a key question is, therefore, whether anonymisation needs to render data incapable of identification by *anyone*? If the answer to this question is 'no', the deletion requirement is essentially defunct, information governance within a particular context becomes far more relevant, and it is far easier to think of anonymity in relative terms. This, in turn, assists with the concept of 'donor anonymity', so it is an important question to consider first.

An initial, if seemingly minor, point to make is that Recital 26 GDPR uses language very similar to that of the former Data Protection Directive 95/46 EC ('the Directive'),[40] but not identical. Recital 26 of the Directive provided:

36 Contractual control, and other means of managing context, are acknowledged on the preceding page of the guidance (page 40), so it is unclear whether the EMA considers the deletion requirement to be engaged in all circumstances.

37 See Rocher, Hendrickx and de Montjoye (n 4).

38 European Commission Directorate-General for Health and Food Safety, 'Question and Answers on the Interplay Between the Clinical Trials Regulation and the General Data Protection Regulation' 7 (2019)<https://ec.europa.eu/health/sites/health/files/files/documents/qa_clinicaltrials_gdpr_en.pdf> accessed 29 January 2020.

39 M Aitken and others, 'Public Responses to the Sharing and Linkage of Health Data for Research Purposes: A Systematic Review and Thematic Synthesis of Qualitative Studies' (2016) 17 BMC Medical Ethics 73.

40 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 (1995) OJ L281/31.

> Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller *or by any other person* to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable . . . (emphasis added)

This is the language on which the 2014 A29WP's Anonymisation guidance was based. The similar, but subtly different, text of Recital 26 GDPR reads:

> The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an 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 (emphasis added).

The difference between 'any other' person and 'another' person may appear trivial, but it goes to the heart of the difference between an objective and a relative understanding of anonymity. It was, in essence, the question at the heart of the *Breyer* judgment: are data personal if they can be identified by 'any' person, or only if they can be identified by the controller and those they are reasonable likely to approach for assistance ('another' person)?

The court considered whether the German government had at their disposal means to identify individuals from IP addresses. The case was referred, and decided, under the Directive, and so the old language of 'any other person' in Recital 26 still applied. Even without the GDPR's clarifying switch to 'another' person, the seemingly objective phrase 'any other' was not taken literally. If an entirely objective view had been taken, the theoretical ability of the Internet Service Provider (ISP) to identify IP addresses would have been sufficient, and only if they deleted their client records could the government data have been out of scope.

Instead, Advocate General Campos Sánchez-Bordona argued that apparently absolute statements within Recital 26 of the Directive should be understood in light of the 'means reasonably likely to be used' benchmark, observing:

> The expression 'means likely reasonably to be used . . . by any other person'. . . could give rise to an interpretation according to which. . . it would be sufficient that any third party might obtain additional data . . ..
>
> 65. That overly strict interpretation would lead, in practice, to the classification as personal data of all kinds of information, no matter how insufficient it is in itself to facilitate the identification of a user. . . .

> 68. Just as recital 26 refers not to any means which may be used by the controller . . . but only to those that it is likely 'reasonably' to use, the legislature must also be understood as referring to '*third parties' who, also in a reasonable manner, may be approached by a controller* seeking to obtain additional data for the purpose of identification. This will not occur when contact with those third parties is, in fact, very costly in human and economic terms, or practically impossible or prohibited by law.[41] (emphasis added)

Paragraph 68 of the Advocate's Opinion, above, was cited approvingly by the CJEU in their subsequent judgment.[42] The IP addresses were still found to be personal, but only because the German government had lawful means to acquire information from the ISP to identify them. The Court apparently heeded the Advocate General's warning that it is impossible to say with certainty that there is no theoretical third party capable of revealing the identity of data subjects. Strict observance of the phrase 'any other person' could, therefore, make a mockery of the 'means reasonably likely to be used' qualification. Under Sanchez-Bordona's argument, therefore, data are personal if they can be identified:

1. by the 'controller' of the information (albeit not necessarily 'data controller' if the data are not personal[43]) and/or
2. by, or with the assistance of, those parties whom said, 'controller' could 'reasonably' approach for aid in identifying individuals.

This pool of parties the controller could reasonably approach for assistance in identification is limited. Cost, both human and economic, should be taken into account—human cost potentially encompassing not only time but also reputational risk. Practical impossibility or prohibition by law are also relevant factors. In instances where the controller is prohibited by contract (or by law[44]) from attempting re-identification, it seems unlikely that they could successfully approach the original data provider and request assistance in identifying data subjects.

Therefore, where such an approach would constitute a breach of contract, or even an attempted criminal offence, it seems much less likely that these parties should be legitimately placed in the camp of third parties who can help identify individuals. As Purtova notes, while legal prohibition cannot mean that individuals are definitely not identifiable,[45] as the risk of illegal acts should not be discounted, it is a factor which

41 *Breyer* (n 8), Opinion of AG Sanchez-Bordona, paras 64–68.

42 ibid, para 46.

43 Under art 4(7) of GDPR, a data controller is the entity which determines the purposes and means of processing personal data. When an entity only controls anonymous data, it is strictly speaking not a data controller according to this definition, but the term 'controller' is still a convenient term to denote their use of the information.

44 For example, it is a criminal offence under s 117 of the UK Data Protection Act 2018 to re-identify de-identified data without the consent of the data controller.

45 Indeed, in light of the above-referenced UK provision, this would mean that all de-identified data in the UK would be automatically considered anonymous for everyone except the data controller and those they authorise to re-identify individuals, thus eliminating the need for security standards and information governance.

may make re-identification *less* reasonably likely.[46] However, when combined with adequate governance measures designed to prevent behaviour which could lead to re-identification (contracts, policies, supervision, professional ethical codes of conduct, as well as legal prohibition), it may be sufficient to say that a controller does not have access to means reasonably likely to be used to identify individuals, either on their own or through reasonable approaches to third parties.

It therefore follows that, even if data could be theoretically identifiable for one party, this does not render them personal for another if they cannot identify people, or reasonably approach others for assistance in identification. This conclusion is hugely helpful for the concept of donor anonymity, where donors and recipients are 'anonymous' vis-à-vis each other, but not to their respective healthcare providers. This calls into question any deletion requirement for anonymisation and opens up considerations of context in assessing the risk of identification. The utility of a more relative account of anonymity is explored in the next section.

IV. DONOR ANONYMITY

This section considers the therapeutic use of donated cells, tissues, and organs. Here, a relative account of anonymity is coherent with the principle of donor anonymity as posited in the legislation. However, unlike healthcare research, the therapeutic relationship with the patient is unlikely to be regulated by a contract in which a patient's use of information can be circumscribed. As such, anonymity through information governance in this context appears to entail the non-disclosure of information as a default, and severely restricted sharing as an exception.

The Tissues and Cells Directive[47] provides a framework for the national law of EU members as regards the donation of tissues and cells. It is supplemented by the directly effective Advanced Therapy Medicinal Product (ATMP) Regulation,[48] which also regulates somatic cell and tissue-engineered therapies, which may use donated cells. It also sits alongside the Organ Donation Directive,[49] which similarly regulates information relating to organ donors. All three of these pieces of legislation refer, in one way or another, to the principle of donor anonymity.[50]

Article 14(1) of the Tissues and Cells Directive, for example, is headed 'Data protection and confidentiality'. It appears to allude to anonymity in the same sense as the term is used under data protection law. It deploys the same terminology of 'rendered anonymous' as used in the former Data Protection Directive and is now used in the GDPR:

46 N Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' (2018) 10 Law, Innovation and Technology 40.

47 Council Directive 2004/23/EC of 31 March 2004 on setting standards of quality and safety for the donation, procurement, testing, processing, preservation, storage and distribution of human tissues and cells [2004] OJ L102/48.

48 Council Regulation (EC) 1394/2007 of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004 [2007] OJ L324/121.

49 Council Directive (EU) 2010/45 of 7 July 2010 on standards of quality and safety of human organs intended for transplantation [2010] OJ L207/14.

50 ibid, Recital 22, also see n 47, Recitals 18 and 29, and n 48, Recitals 15 and 19.

> Member States shall take all necessary measures to ensure that all data, including genetic information, collated within the scope of this Directive *and to which third parties have access*, have been *rendered anonymous* so that neither donors nor recipients remain identifiable. (emphasis added)

It seems clear that the term 'rendered anonymous' is intended to have the same meaning as in data protection legislation, as it comes in an Article headed 'Data protection' which uses the same language. In only requiring data 'to which third parties have access' to be rendered anonymous, it also seems to draw on a relative interpretation of anonymity in which data can be 'rendered anonymous' for a third party but remain personal data for the controller. As Article 14 goes on to say:

> 2. For that purpose, they shall ensure that:
> (a) data security measures are in place, as well as safeguards against any unauthorised data additions, deletions or modifications to donor files or deferral records, and transfer of information;
> (b) procedures are in place to resolve data discrepancies; and
> (c) no unauthorised disclosure of information occurs, *whilst guaranteeing the traceability of donations.*'
> 3. Member States shall take all necessary measures to ensure that the identity of the recipient(s) is not disclosed to the donor or his family and vice versa, without prejudice to legislation in force in Member States on the conditions for disclosure, notably in the case of gametes donation (emphasis added).

2(c) is particularly interesting, as it makes it clear that when information is provided to third parties, the act of rendering data anonymous cannot include the deletion of the original data, as the traceability of the donor and recipient must be maintained. This is clearly incompatible with the deletion/aggregation requirement in the A29WP's guidance. The guidance would require data to be rendered anonymous for the controller, even though they are authorised to hold this confidential information, and their ability to identify subjects does not necessarily equate to that of a third party. This requirement would be particularly impracticable in the context of tissue and cell donors and recipients, who must remain traceable at least by those authorised to hold their confidential information. Even though there is no question of the donor or recipient's medical safety being compromised by the deletion of their own healthcare records, the word 'anonymous' is nonetheless used in the Tissues and Cells Directive. This suggests that the two parties do not need to be anonymous for everyone, to be anonymous to each other.

It could be argued that, as the A29WP was specifically turning its mind to anonymisation, and the drafters of the Cells and Tissues Directive were not, the 2014 anonymisation guidance should prevail, and the Cells and Tissues Directive should be amended accordingly. Viewed from this perspective, the word 'anonymous' should be understood in a loose sense within the Cells and Tissues Directive and should perhaps be replaced with the word 'pseudonymous' instead.

However, there are two problems with contention. First, if the term 'anonymous' or 'anonymity' were removed from the Cells and Tissues Directive, it would be difficult to ensure that donors were protected to the desired standard. If the word 'anonymous' were replaced with 'pseudonymised' or 'de-identified', there would be no accompanying requirement for the information disclosed to be anything other than personal data, that is, capable of identifying individuals by means reasonably likely to be used. The protection offered to subjects would thus be weaker if all the Directive required was for data providers to render them *less* identifiable, without any thought as to whether they are still reasonably likely to be identified by the recipient. This level of identity protection is unique to the term 'anonymous'; without this standard, subjects cannot be offered a reasonable level of confidence that their identity will not be revealed to third parties.

Secondly, as observed at the beginning of this subsection, the Tissues and Cells Directive deliberately uses the language of rendering data anonymous and was thus evidently not intended to be used in a loose sense. A more compelling interpretation is that the Tissues and Cells Directive, like the Organ Donation Directive and the ATMP Regulation, refers to donor anonymity meaning these individuals should not be identified by each other, or any other unauthorised parties, by any means reasonably likely to be used. In other words, information disclosed to a non-authorised party should not reveal their identity by means reasonably likely to be used, but they do not need to be (and cannot be) anonymous for everyone. Recital 29 of the Tissues and Cells Directive, therefore, appears to use the word 'anonymity' in a relative (or 'subjective') sense:

> As a general principle, the identity of the recipient(s) should not be disclosed to the donor or his/her family and vice versa, without prejudice to legislation in force in Member States on the conditions of disclosure, which could authorise in exceptional cases, notably in the case of gametes donation, the lifting of donor anonymity.

The meaning of 'anonymity' must correlate to the similar instruction in Recital 22 of the Organ Donation Directive:

> As a general principle, the identity of the recipient(s) should not be disclosed to the donor or the donor's family or vice versa, without prejudice to legislation in force in Member States which, under specific conditions, might allow such information to be made available to donors or donors' families and organ recipients.

Changing this text to refer to 'donor pseudonymity' would not prevent the disclosure of information reasonably likely to identify these individuals. Requiring raw data to be deleted, or aggregated, to achieve anonymity for all parties would be clearly contrary to the requirement of traceability. This is strongly supportive of a relative account of anonymity, without which anonymity in the context of traceable donations would be impossible.

In short, the principle of donor anonymity illustrates the limited, relative sense in which the word 'anonymous' is used in EU legislation beyond the GDPR. It is a

relatively straightforward proposition where donors and recipients are given little or no information about each other. As explored further in the next section, anonymity is more challenging when information is publicly accessible and, therefore, no longer subject to information governance.

V. CLINICAL TRIALS REGULATION

The CTR[51] came into force in 2014 with the aim of simplifying and harmonising the governance of clinical trials in the EU.[52] A core aspect of the CTR is the creation of a central portal through which applications for authorisation of clinical trials, and subsequent mandatory information, can be sent by sponsors to the relevant body within each Member State.[53]

Information submitted through the portal will be stored in a publicly accessible database, termed 'the EU database'.[54] This EU database is intended to include mandatorily submitted information about clinical trials, but not the personal data of trial participants. Clinical Trial participants, therefore, provide another case study of relative anonymity within health law, as they should be identifiable within trials but not to the world at large.

A. Information Submitted to the EMA

The CTR provides that annual safety reports of investigational medicines should be submitted to the EMA and stipulates:

> 3. The annual report referred to in paragraph 1 shall only contain aggregate and anonymised data.[55]

This is striking in its use of the term 'aggregate and anonymised' data. In contrast to 'donor anonymity', it is less clear what is meant by 'anonymised' in this context. It could mean that aggregate and anonymised data are different types of information, or that anonymisation requires aggregation, depending on whether the word 'and' is conjunctive or disjunctive. Under a relative account of anonymity, however, it would be possible for individual-level information to be submitted to the EMA, where necessary, while still qualifying as anonymised.

While individual-level information should obviously be kept to the minimum necessary for the safety report, a sponsor providing limited individual-level information (stripped of direct identifiers) *to* a public authority is in a very different position to private citizens receiving donor information *from* a public authority. The focus in the CJEU's *Breyer* judgment on the lawful means that the German government had at their disposal to obtain additional information to identify individuals was key to their finding that the dynamic IP addresses of those visiting government-controlled websites were personal data. The converse inference is that, if the German government

51 Council Regulation (EU) 536/2014 of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC [2014] OJ L158/1 (hereafter cited as 'CTR').

52 ibid, Recital 4

53 ibid, art 80.

54 ibid, art 81.

55 ibid, art 43.

had no lawful means of identification, the IP addresses would not have been personal data in their hands. While it should not be inferred as a general rule that all illegal re-identification is, therefore, not reasonably likely,[56] there is a more compelling argument that public authorities are vested with a level of trust to behave within the law, which goes above and beyond the expectations placed on ordinary citizens and corporations. The safeguards needed to prevent re-identification could thereforebe lighter where such institutional trust would reasonably be deemed appropriate.

It is naturally not realistic to expect sponsors providing clinical safety reports to demand the right to audit the EMA to ensure that their data subjects are not re-identified, as might be the case in medical research collaboration (see Section II). But, at the same time, courts may well judge it reasonable for a sponsor to trust the EMA not to break the law. It could, therefore, be reasonable to submit minimised, individual-level information to the EMA where necessary for a safety report, on the understanding that the Agency is not authorised to re-identify people from this minimised information. This trust is arguably as important a safeguard as any data modification technique. The same argument, however, does not work when individual-level information is made available to the public at large, which could be accessed by any range of trustworthy or non-trustworthy actors.

B. Information Made Publicly Accessible

The compromise which appears to emerge from the above is that it could be possible for information to be provided to the EMA on an anonymous basis. However, information made publicly accessible in the EU database is more difficult to anonymise reliably as the perspective of the potential recipients of that information is more difficult to anticipate.

The CTR appears to accommodate this, as Article 81(7) CTR stipulates that 'No personal data of subjects shall be publicly accessible.' By implication, although the EU database will be publicly accessible by default, where personal data of trial participants could be exposed by the release of information, that data should be excluded from the publicly accessible aspects of the register. For example, Recital 67 GDPR suggests that clinical study reports should be included in the EU database:

> The EU database should be publicly accessible and data should be presented in an easily searchable format, with related data and documents linked together by the EU trial number and with hyperlinks, for example linking together the summary, the layperson's summary, the protocol and *the clinical study report of one clinical trial*. (emphasis added)

Examination of the Medicinal Products Directive[57] (which specifies the contents of clinical study reports) and the EMA's own guidance[58] suggests that clinical study

56 Purtova (n 46).

57 Council Directive 2001/83/EC of 6 November 2001 on the Community code relating to medicinal products for human use [2001] OJ L311/67, Annex 1, Module 5.

58 European Medicines Agency, 'Note for Guidance on Structure and Content of Clinical Study Reports' (2006) <https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-3-structure-content-clinical-study-reports-step-5_en.pdf> accessed 19 July 2019.

reports should contain individual-level information, for example, including information about any participants prematurely withdrawn from a clinical trial. The EMA notes that, in respect of each participant discontinued after enrolment (who should be identified by a patient identifier):

> It may also be useful to include other information, such as critical demographic data (e.g. age, sex, race), concomitant medication, and the major response variable(s) at termination.[59]

This would be problematic if this detailed information about participants was made available to the public at large, given the requirement that no participant personal data should be recorded in the public database.[60] It is unlikely to be appropriate to call such individual-level information anonymous, unless a strong argument could be made that the information is insufficiently unique for anyone—including the subject, their family, or doctor—to identify them. Even this argument is susceptible to miscalculation, however, and it would be preferable that 'anonymous' data were only shared with a party subject to strict re-identification controls, including to a public authority who should be trusted to behave within the law, and not illegally re-identify subjects. To return to the test discussed in Section III, it is no longer possible to say who the 'controller' is, and whom they might reasonably approach, if the said controller is a theoretical 'anyone' within the general public, and so without information governance the relative account of anonymity no longer functions.

This supports the case for submission of any necessary individual-level information relating to clinical trial subjects to the EMA, but not publishing such information on an open, 'anonymised' basis. Public disclosure without governance controls, solely reliant on data modification techniques, risks perpetuating the long-standing trend of re-identification from publicly released information.[61] This, in turn, risks fuelling criticism of anonymity by mischaracterising it as a set of techniques rather than a legal standard which should reasonably be met through a more holistic assessment and management of information use.

The next section addresses criticisms of anonymisation and considers the extent to which these stem from the way it has been perceived as a potentially unsuccessful set of techniques. This ultimately culminates in a defence of the continuing utility of the standard, as an important benchmark as to when subject identities are (not) sufficiently protected.

VI. CRITICISMS OF ANONYMISATION

In light of the above exploration of the term 'anonymous' in data protection and other EU law, anonymisation can be described as the *result* of practices which:

59 ibid

60 CTR (n 51), art 81(7) and Recital 67.

61 Ohm (n 16); Rocher, Hendrickx and de Montjoye (n 4).

- ensure through governance as well as data modification that information revealed to a third party does not relate to any natural person they can identify (by means reasonably likely to be used);
- as such, mean that the confidentiality of the information has been protected; and
- mean that, strictly speaking, the recipient does not need to comply with the GDPR. However, it would be difficult to argue that reasonable means of identification have been excluded if the governance does not involve significant restrictions on the use of individual-level data, which may in practice closely mirror GDPR requirements.

Correspondingly, anonymisation does *not*:

- enable the recipient to do whatever they like with the data, as this would be contrary to any adequate controls against re-identification;
- mean that the data controller is not processing personal data in making the disclosure—the data are still identifiable *for them* at the point of disclosure. The disclosure should, therefore, still comply with the GDPR, including provisions relating to transparency; and
- resolve the ambiguity as to whether the discloser remains a data controller of 'anonymised' data, where they have set strict terms for its use.[62]

In short, anonymisation is a means of preserving confidentiality by protecting individual identities to a reasonable standard. It cannot be achieved without multi-faceted attempts to limit what is done with individual-level data. As such, it does not represent a more liberal regime of data processing, but merely one in which information can be disclosed to a particular recipient without specific consent or breach of the duty of confidence.

In light of this proposed characterisation, this section briefly addresses some of the long-standing criticisms of anonymisation in order to clarify the extent to which this characterisation of anonymisation addresses these concerns.

To start with, it should be acknowledged that even a relative, governance-based approach to anonymisation will not prevent what have been characterised as group harms. Floridi, for example, has argued that it is a 'very dangerous fallacy to think that if we protect personal data that identify individuals, the protection of the groups will take care of itself'.[63] Likewise, the protection of anonymised individual-level data does not prevent inferences drawn at group level being reapplied to the same or other individuals in a harmful way. Taylor highlights genetic data as an example of information which relates to the privacy of whole minority ethnic groups and for which individual-level consent or identity protection offers inadequate redress.[64] Inferences drawn

62 UK case law suggests this is not the case, and the discloser would only be the controller for their own internal processing, with subsequent processing by the recipient outside the scope of data protection law, for example, *Common Services Agency (CSA) v Scottish Information Commissioner* [2008] UKHL 47. This may be debatable, however, if the discloser imposes conditions on the use of what are (for them) personal data.

63 L Floridi, 'Open Data, Data Protection, and Group Privacy' (2014) 27 Philosophy & Technology 1.

64 M Taylor, *Genetic Data and the Law: A Critical Perspective on Privacy Protection* (1st edn, CUP 2012) 150–51.

from personal or anonymised data may be easily detached from the individual-level information from which they were derived and formulated at a generalised level, but nonetheless reapplied to individuals when they are stratified into, for example, categories supporting decisions relating to health insurance or even prescribing practices.[65] It has, therefore, been argued with some justification that inferential data pose the greatest risks in terms of privacy and discrimination but are offered the least protection under data protection law.[66]

To the extent that measures to prevent re-identification mirror the protections required by the GDPR, and anonymisation thus mirrors data protection, anonymisation will also replicate data protection's weaknesses and limitations. There is the same focus on individual-level rights, and an inability to govern the intellectual consequences of data processing. This is why exploration of law against data-driven discrimination,[67] discussion of more systemic oversight of Big Data in health,[68] or even Harm Mitigation Bodies to scrutinise downstream effects of data use[69] are of increasing value as use of advanced analytics increases, and legal anonymity can only be a solution to a limited set of problems. The Anonymisation Decision-Making Framework rightly addresses this by incorporating ethical reflection as part of the anonymisation process,[70] which could encompass debate on potential group harms, but this goes above and beyond what is required to manage identifiability to the legal standard of anonymity, and cannot be guaranteed as an aspect of legal compliance.

A governance-based approach does, however, help to address the concern that anonymisation creates a Wild West of Big Data in which artificial manipulations of data can enable any and all subsequent uses. It prevents some of the danger posed by data-centric, governance-light 'release and forget' models in which assessing identifiability from the perspective of the discloser creates a false sense of objectivity, as though making data less identifiable *for them* will necessarily make them non-identifiable for everyone else. Given the prominence of such modification-based approaches, it is not surprising that anonymisation has been described as a technical solution, rather than an ethical one.[71] Neither is it surprising that some see the distinction between personal/anonymous data as being determined at the point of collection (or disclosure) and then ignored as identification risk fluctuates unchecked within subsequent usage.[72] Andrew and Baker, for example, argue that the law lacks a vital understanding

65 M Ravindranath, 'How Your Health Information Is Sold and Turned into "Risk Scores"' (2019) <https://www.politico.com/amp/story/2019/02/03/health-risk-scores-opioid-abuse-1139978?__twitter_impression=true> accessed 6 April 2020.

66 S Watcher and B Mittelstadt, 'A Right to Reasonable Inferences: Re-thinking Data Protection Law' (2019) 2 Columbia Business Law Review 443.

67 Although such laws may also have significant limitations—see W Nicholson Price II and I Glenn Cohen, 'Privacy in the Age of Medical Big Data' (2019) 25 Nature Medicine 37.

68 E Vayena and A Blasimme, 'Health Research with Big Data: Time for Systemic Oversight' (2018) 46 The Journal of Law, Medicine & Ethics 119.

69 A McMahon, A Buyx and B Prainsack 'Big Data Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond' (2019) 28 Medical Law Review 155. .

70 See n 29.

71 G Laurie and L Stevens, 'Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom' (2016) 43 Journal Law and Society 3.

72 S Watcher, 'Data Protection in the Age of Big Data' (2019) 2 Nature Electronics 6.

of contemporary big data practices, and this may be attributable to an overzealous trust in technical solutions.[73]

These are all indeed legitimate criticisms if modified, individual-level data are disclosed to one or many parties without assessment and control of re-identification risk in the new environment(s), and this process is referred to as 'anonymisation'. It is a central argument of this article, however, that such an act cannot be seen as rendering data anonymous. Whether there exists a means reasonably likely to identify someone is inevitably a function of the context in which those data are situated. As stated in a UK House of Lords' judgment on the release of modified medical information relating to children:

> Whether or not the individuals are identifiable from the barnardised data is a question of fact, the answer to which may vary from situation to situation and, indeed, from individual to individual.[74]

If there is an overzealous trust in technical solutions, this is not the fault of the law but of those seeking to apply it. The law has set a reasonable standard, to achieve the explicitly limited purposes of privacy and data protection, but its implementation relies on an adequate assessment of identification risk, which is a question of fact. Those who ignore context, and the extent to which such context can be adequately reviewed and controlled through information governance, do the law a disservice.[75] Technical solutions cannot be a substitute for an analysis of circumstance, and the law does not suggest they should be. To return to the Advocate General's opinion in the *Breyer*, factors which should be taken into account in determining whether data are personal include consideration of third parties a controller may approach, human and economic cost, practical (im)possibility, and prohibition by law.[76] Neither the GDPR itself nor legal authority determined under the Data Protection Directive promotes a context-free approach to anonymisation.

A. Relational Anonymity?

It is understandable that discussion of the reuse of health-related information raises anxiety about severing the relationship between patients and their data,[77] and anonymisation (at least in its absolute form) can be such a severance between information and the rights formerly associated with it. However, as Ballantyne has argued, a better response to this risk is not to entrench connection through ideas of individual ownership, but instead develop flexible models to reconnect patients with their data,[78] with emphasis on consent, transparency, and engagement.

A relative understanding of anonymity could fit within such a flexible model. If data are personal for the organisation collecting the information, but not for the third

73 Andrew and Baker (n 3).

74 *Common Services Agency* (n 62) para 87.

75 Again, see Elliot and others on the distinction between failures in privacy law and practice (n 17).

76 See n 41.

77 A Ballantyne, 'How Should We Think about Clinical Data Ownership?' (2020) Journal of Medical Ethics <10.1136/medethics-2018-105340> accessed 7 February 2020.

78 ibid.

parties with whom it is shared on a controlled access basis, the collector is still subject to the GDPR and subjects can enforce their rights against this entity, without exposure to identification risk from third parties.

Furthermore, it is important to remember that the rights of privacy, confidentiality, and data protection we fear severing are themselves contextual, and even relational. Taylor and Wilson have demonstrated that the touchstone for a (medical) duty of confidence is now a reasonable expectation of privacy, which (like anonymity) is determined based on all circumstances of the case.[79] They argue that careful respect for autonomy is vital for compliance with these reasonable expectations, although in practice this autonomy may be of a more collective, relational nature[80] that could also be accommodated within 'flexible' transparent models of data use.

A useful illustration of the shift towards relational privacy comes from the Court of Appeal of England and Wales in *R (W, X, Y and Z) v Secretary of State for Health and Secretary of State for the Home Department*[81]:

> the question whether there is a reasonable expectation of privacy is a broad one which takes account of all the circumstances of the case. We do not see how overseas visitors who, before they are treated in an NHS hospital, are made aware of the fact that, if they incur charges in excess of £1,000 and do not pay them within 3 months, the Information may be passed to the Secretary of State for onward transmission to the Home Office for the stated immigration purpose can have any, still less any reasonable, expectation that the Information will not be transmitted in precisely that way. They will, however, have a reasonable expectation of privacy in relation to the Information vis-à- vis anyone else.

There is an alignment between this characterisation of privacy as arising vis-à-vis some people, but not others, and the idea of anonymity as existing vis-à-vis some people but not others, depending on a broad range of factors. An individual may be reasonably likely to be identified by someone, but not have a reasonable expectation of privacy in relation to that party. Conversely, they may have a reasonable expectation of privacy in relation to someone, but not be reasonably likely to be identified by them. In other cases, reasonable expectations of both identification and privacy will apply and data protection, confidentiality, and privacy rights will all be engaged.

The contextuality of identifiability and privacy rights is complex, but congruent. Understood in this way, anonymity does not draw a bright, severing line between people and their data protection rights. It is, I suggest, one of at least two key context-specific distinctions that run through data which helps to determine what kind of rights are engaged. This helps to locate relative anonymity in a broader context of information governance law, which adds weight to the characterisation as initially discussed in terms of cell and tissue donation.

79 MJ Taylor and J Wilson, 'Reasonable Expectations of Privacy and Disclosure of Health Data' (2019) 27 Medical Law Review 3.

80 E Dove and others, 'Beyond Individualism: Is There a Place for Relational Autonomy in Clinical Practice and Research?' (2017) 12 Clinical Ethics 3.

81 *R (on the application of W, X, Y and Z) v Secretary of State for Health and Secretary of State for the Home Department, the British Medical Association* [2015] EWCA Civ 1034, [44].

VII. CONCLUSION

This article has advocated a relative, governance-based understanding of anonymity. Anonymity has been shown to be a legal standard, which requires the elimination of reasonable means of identification. I have argued that, where 'anonymisation' fails, this is a failure to meet the standard of anonymity, and not a failure of the standard itself.

The relative account has been explored via a number of pieces of legislation. We have seen that Recital 26 GDPR has been interpreted by the CJEU with some relativity, and not as meaning that data must be unidentifiable for every theoretical beholder to be 'rendered anonymous'. Crucially, I have argued that data do not have to be anonymous for a discloser to be anonymous for a recipient, as long as the recipient could not reasonably approach them for assistance in identifying people (and do not have other reasonable means of identification). The ability of a discloser to identify subjects is, I have suggested, a poor benchmark to judge identifiability for everyone else. It is often impossible for them to delete their own raw data, but this should not prevent them from sharing information to the standard of anonymity and protecting subjects from identification by a third party.

This relative perspective has been assisted by an analysis of donor anonymity and the anonymity of clinical trial subjects under the CTR. In both contexts, it appears that some individual-level information must be shared, even when it might well be identifiable for the discloser. I have argued that this is compatible with anonymity as seen from a relative perspective. However, I have also argued that anonymity should be assessed with reference to governance, and the mechanisms in place to prevent re-identifying behaviour, which points against the publication of individual-level information on an 'anonymous' basis.

Despite its limitations, the value of anonymisation lies in its capacity to assure subjects that they are not reasonably likely to be identified, and their privacy and confidentiality have, therefore, been preserved. It does not maintain trust in the concept to suggest that anonymisation consists of techniques which open the door to unregulated uses of data, when in fact the maintenance of anonymity requires an equivalent standard of protection. While this may not necessarily prove an easier alternative to data protection law, its purpose is to protect subjects' identities and, thus, their rights to privacy and confidentiality, and not to serve controller convenience.

This account of anonymity, if accepted, naturally has implications outside the medical context. For example, the UK has recently established a system for reuse of non-medical administrative data for research under the Digital Economy Act 2017. While de-identification of data is an initial step in minimising identification risk, oversight, training, and accreditation of all bodies involved in processing the information helps to eliminate likely means of identification, at least for the ultimate recipients of such information. Accredited researchers, processors, and peer-reviewers must agree to be listed in a public register for transparency purposes, along with a summary of the approved project.[82] Data access agreements are supplemented by a Code of Practice for

82 UK Statistics Authority, 'List of Accredited Researchers and Research Projects under the Research Strand of the Digital Economy Act' (2019) <https://www.statisticsauthority.gov.uk/about-the-authority/better-useof data-statistics-and-research/betterdataaccess-research/better-use-of-data/list-of-accredited-researchers-and-research-projects-under-the-research-strand-of-the-digital-economy-act/> accessed 27 February 2020.

data sharing, as well as a long list of counts by which researchers can lose their accreditation, including negligently facilitating the identification of individuals in the data.[83] The scrutiny, public profile, training, and legal obligations compounded within this statutory structure provide a helpful example of governance measures which help to minimise identification risk and support relative anonymity. It is suggested that approaches such as these could be more important in a move towards a conceptualisation of anonymity grounded in information governance.

Conflict of interest statement. None declared.

83 UK Statistics Authority, 'Research Code of Practice and Accreditation Criteria' (2018) <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/08/COP_Research-and-Accreditation_A4.pdf> accessed 27 February 2020.

Publication 3

Publication Title: Identity, Profiles and Pseudonyms in the Digital Environment

Short Title: ‘Profiling Chapter’

Page Number of Thesis: 57

The following pages incorporate a PDF of this book chapter. It is an extract of a PDF of the whole book, *The Boundaries of Data* (van Der Sloot and van Schendel, eds) which has been published online on an open-access basis. The formatting, pagination and footnote numbering of the publication have been retained, and thus the numbering is self-contained.

The pagination sequence of this thesis will then resume at the end of the publication

14. Identity, Profiles and Pseudonyms in the Digital Environment

Miranda Mourby[1] *& Elaine Mackey*

Abstract
The boundaries of personal data are determined by the concept of 'identity'. Personal data, as defined under the GDPR, is information relating to an identified or identifiable natural person. In this chapter, we argue that the informational 'identity' of an identified/identifiable person is characterised by the potential for privacy impact. Our informational identity is, in essence, the sum of all the information which can impact our rights. We use profiles and pseudonyms as an illustration of this definition. Profiles permit scrutiny of an individual – and thus 'identify' them through the intrinsic privacy impact of this evaluation. Pseudonyms alone do not allow individuals to be evaluated, which is why they are not, in and of themselves, personal data.

Keywords: identity; pseudonymisation; profiling; anonymisation; personal data

1. Introduction

What is an identification? Some information is deemed sufficiently 'us' to warrant legal protection, but this category of information shifts all the time, and the logic underpinning these shifting parameters is far from explicit. The idea of 'identity' determines the scope of data protection law in the EU, which safeguards the rights of 'identified' and 'identifiable' individuals. Without understanding when a person is – or might be – 'identified', we cannot be sure when these rights arise.

1 Miranda Mourby would like to acknowledge support from the EU-STANDS4PM consortium (www.eustands4pm.eu) that was funded by the European Union Horizon2020 framework programme of the European Commission under Grant Agreement #825843. She is also grateful to the School of Law at the University of Sheffield, whose funding supported this work in part.

Van der Sloot & van Schendel. *The Boundaries of Data*. Amsterdam: Amsterdam University Press, 2024. DOI: 10.5117/9789463729192_CH14

This chapter clarifies the concept of 'identity' in EU and associated national data protection law by flipping conventional wisdom on its head. It is often asserted that privacy and data protection rights arise when an individual is or can be identified. But without a clear understanding of what it means to be 'identified', this statement is not particularly meaningful. As the growth of the online infosphere increasingly detaches identity from traditional 'real-world' signifiers, the time may have come to recognise that an individual is instead 'identified' when information engages their rights to privacy and/or data protection. As profiling is thought to engage privacy and data protection rights and is proliferating within the Big Data environment (de Hert & Lammerant, 2016), it is a useful touchstone in understanding identification in digital information.

This chapter therefore attempts to delineate the contours of 'identity' in data protection law by exploring two associated concepts: profiling and pseudonymisation. We have selected these concepts because they are respectively associated with *direct* and *indirect* identification. We suggest that the parameters of 'direct' identification – information that is, in and of itself, an identification with nothing further required – help to reveal the nature of an identity in data protection law. The UK is used as a particular case study because it has, in its post-Brexit modification of the EU General Data Protection Regulation (GDPR), introduced the concepts of direct and indirect identification into a statutory definition of identifiable individuals, which adds precision to the definition that can be inferred at EU level.

The concepts of pseudonymisation and profiling under the GDPR are therefore worth unpacking because they help illustrate the circumstances in which identification takes place in the online infosphere. In the absence of a definition of 'identified' or 'identifiable' individuals in the EU Regulation itself, these subsidiary concepts provide contrasting definitions of a directly identifying 'profile' (which engages an individual's rights through evaluation of their personal characteristics) with a 'pseudonym' (which also uniquely represents people but does not permit analysis or scrutiny of them as individual subjects without further information). The 'unique' nature of the pseudonym may only be a particular variation in a hashing code; it does not signify any immediately discernible personal information. Put simply, therefore: if a profile alone is an identification, and a pseudonym alone is not, the contrast between the two helps us explain what is and is not an identity in online information.

Ultimately, we suggest that the defining feature of 'identity' in data is the capacity of information to interfere with individuals' privacy and data protection rights. As profiling data permit scrutiny of individuals in a way that pseudonymised data should not, this distinction between the two

concepts provides a useful illustration of the difference this capacity of interference makes in practice.

2. Identity in Data Protection Law

As Sullivan (2011) emphasises, it is important to discern the meaning attributed to the concept of 'identity' in a particular legal context:

> Identity has traditionally been a nebulous notion and in referring to 'identity' without defining it, much of the legal literature in this area lacks precision. It gives the impression that 'identity is identity' whereas the constitution, function and nature of identity depends on context … it is important to differentiate the 'purely legal relations' from other non-legal conceptions. (p. 6)

In order to delineate the meaning of identity in the context of data protection law, it is necessary to grapple with the GDPR's usage of the terms 'identified', 'identifiable' and 'identifier.' These occur in the definition of personal data in the GDPR:

> 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. (GDPR, Article 4[1])

It is easy to lose one's bearings within a definition so densely packed with the terms identified, identifiable, identifier and identity. Interestingly, while the term 'identifiable' is elaborated upon as meaning someone who 'can be identified', the word 'identified' itself is not explained, leaving an ultimate ambiguity as to what 'identity' means for the purposes of the GDPR. The list of 'identifiers' is perhaps a clue, but these pieces of information appear only to refer to *means* of identification and not identification itself. As the UK Information Commissioner's Office (2020) clarifies, 'whether any potential identifier actually identifies an individual depends on the context.' For example, 'a person who enjoys the theatre' may be an aspect of cultural identity, but without further information to link this no doubt scintillating insight into one particular person, it is no more identification than it is trope, fiction or hypothesis.

We have suggested that two types of personal data can be established within the GDPR:

– information that is, in and of itself, identification (relating to an 'identified' individual);
– information that can be linked indirectly to an identified individual, including pseudonymised data, which is information on an 'identifiable' individual.

In order to answer the question of what identification is, we are principally concerned with the first type of personal data – information that is *in itself* an identification. The latter category is essentially a secondary subset of personal data, caught by the regulation if they can be linked to information that either in combination or in itself constitutes identification. The core question, therefore, is what quality or qualities of data render information an identification.

We will answer this question of 'what is identification?' by relating to direct identification, i.e. information relating to *identified* individuals. Within privacy and data protection, data that are characterised as 'personal' – and therefore as linking to individuals' 'identity' – tend to be information with sufficiently close association to an individual to justify their 'stake' in the information. As Laurie states in the context of genetic data, 'individuals have an interest in this information because it relates to them and can affect their lives' (Laurie, 2002).

2.1. Facial Images as Direct Identification

A UK case that illustrates this association with identification and the idea of a personal stake in information is the High Court judgment in *Bridges v. South Wales Police*, which was believed to be the first time any court in the world had considered the use of automated facial recognition software (AFR). The claim for, inter alia, infringement of data protection legislation was brought by Edward Bridges with the support of the campaigning organisation Liberty.

In brief, *Bridges* concerned the collection of facial images by police at rugby matches for the purposes of AFR. It was argued in submissions that the police would require further powers to match the facial images to individuals in order for them to constitute personal data (per *Breyer*). In other words, the images were not an identification in and of themselves, and 'identifiability' would only be triggered with the presence of an additional means reasonably likely to be used to identify people.

The Court rejected this argument, however, on the basis that the images were an identification in and of themselves:

> Where the data in issue is biometric facial data, we see no need for the analysis adopted by the CJEU in Breyer (in the context of information comprising dynamic IP addresses). Whether or not such information is personal data may be open to debate, as is apparent from the judgment in Vidal-Hall [2016] QB 1003. However, the biometric facial data in issue in this case is *qualitatively different* and clearly does comprise personal data because, *per se*, it permits immediate identification of a person. (R. [on the application of Bridges], 2020; emphasis added)

The phrase 'immediate identification' makes it clear that an image of a face is an identification in and of itself, having the 'quality' of being identity per se. This is reminiscent of Sullivan's description (cited above) of the 'identity is identity' mentality. Although the reasons for this are not elaborated upon, it seems overwhelmingly contextually likely that the Court bore the civil liberty implications mentioned above in mind, meaning that the location of the information within the regulatory framework of privacy and data protection was a pressing concern in this determination. The risks revealed by the evolution of AFR thus make a compelling argument for consideration of images of faces as an identification, and thus an identification in the eyes of data protection law.

2.2. IP Addresses as (In)Direct Identification

IP addresses, on the other hand, are not as straightforward a proposition. An IP address alone is not necessarily an identification because it does not create sufficient potential for consequence for, or inference about, the user of the related device, but an IP address combined with browsing history data across a number of websites *is* generally held to be an identification because it creates a profile. Evidence for this argument can be found in Recital 30 GDPR:

> 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)

This recital seems to draw a reasonably clear distinction between potential identifiers (such as an IP address) and the combination of information that

profiles an individual, adding up to enough usable information to constitute an actual identification.

Further illustration of how IP addresses can fail to meet the standard of direct identification comes from the 2016 judgment Case C-582/14 of the Court of Justice of the European Union in *Patrick Breyer v Bundesrepublik Deutschland*, which we will refer to as the *Breyer* judgment.

In the *Breyer* case, the German government collected information in case its websites came under attack and it was necessary to identify the perpetrators:

> With the aim of preventing attacks and making it possible to prosecute 'pirates', most of those websites store information on all access operations in logfiles. The information retained in the logfiles after those sites have been accessed include the name of the web page or file to which access was sought, the terms entered in the search fields, the time of access, the quantity of data transferred, an indication of whether access was successful, and the IP address of the computer from which access was sought. (CJEU, 2016, para. 14)

These retained IP addresses had no immediate privacy consequences for the associated individuals unless the German government took additional steps to build a picture of these people. It was confirmed at paragraph 38 of the judgment that the dynamic IP addresses were not personal data in and of themselves:

> In that connection, it must be noted, first of all, that it is common ground that 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. (CJEU, 206, para. 38)

3. Pseudonyms and Profiles

The terms 'pseudonyms' and 'profiles' are used in this chapter to refer to the end products of GDPR pseudonymisation and profiling respectively. While these terms may, in other contexts, both refer to representations of individuals that fall short of an identification (e.g. a psychological 'profile' of a criminal suspect who sends letters under a 'pseudonym' but has yet to be

identified), in the context of EU data protection law, they denote different levels of identifiability.

A 'pseudonym' is traditionally defined as an alternative to one's 'real' identity, for example as a 'false or fictitious name, esp. one assumed by an author; an alias' (Oxford University Press, 2007). In the context of the GDPR, personal data that have undergone pseudonymisation are associated with an 'alias' or something falling short of an actual identification. The data thus requiring additional information to be linked back to the 'real' identity of the natural person:

> 'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. (Article 4[5] GDPR)

A profile, by contrast, permits the evaluation of personal characteristics under its definition in 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.

This automated evaluation of personal characteristics is, we suggest, sufficient intrusion into privacy and data protection rights to constitute an identification in and of itself, even if there are no other consequences for the data subject. For example, a profile of an individual's online behaviour is likely to involve novel inferences about that person, which are of value for commercial exploitation, which then steps over the boundary of anonymous, unobserved browsing even before any attempt to 'reach' or affect the individual is made. The use of profiling in the digital environment therefore illustrates the underlying logic of identification: where there is intrusion, there is identification, even if the digital profile bears questionable resemblance to someone's 'real' identity.

Table 14.1 attempts a summary of how we distinguish the GDPR terms 'profiling' and 'pseudonymisation':

<table>
<tr><th></th><th>Pseudonymisation</th><th>Profiling</th></tr>
<tr><td>Definition under Article 4 GDPR</td><td>The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.</td><td>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.</td></tr>
<tr><td>Information associated with the process</td><td>A pseudonym plus a string of information</td><td>A string of information that allows an evaluation of personal characteristics</td></tr>
<tr><td>Example</td><td><table><tr><td>Masked IP address</td><td>Age in bands of ten years</td><td>Gender</td><td>Employed [y/n/]</td><td>Industry of employ-ment</td><td>Number of times X website visited last 12 months</td></tr></table></td><td><table><tr><td>IP address</td><td>Website address</td><td>Date visited</td><td>Time + Duration of visit</td><td>Products viewed</td><td>Products bought</td></tr></table><table><tr><td>Masked IP address</td><td>Website address</td><td>Number of times Website visited last 12 months</td><td>Most popular product viewed last 12 months</td><td>Most popular product bought in the last 12 months</td></tr></table></td></tr>
<tr><td>How the processes differ</td><td>– A process to mask identity
– A process that reduces detail in the personal information, leaving the substance of the data untouched
– Aims to prevent direct identification</td><td>– A process to create an identity (may not match real world identity)
– A process to create new information, 'a profile', from the underlying original data
– Aims to evaluate the individual</td></tr>
<tr><td>Purpose of the process</td><td>To dissociate individuals from information. It enables the exploration of patterns within data whilst preventing direct identification.</td><td>To associate individuals with information. It enables the targeting of natural people in pursuit of a service, product or message with the potential to 'reach' the individual.</td></tr>
</table>

Table 14.1

3.1. Profiles as Direct Identification: IAB Europe

An important example of profiling is the 'Transparency and Consent', or 'TC String', generated by consent management platforms to record the consent preferences of visitors to websites regarding the use of their data.

This 'TC String' was considered in the judgment of Case DOS-2019-01377 before the Litigation Chamber of the Belgian Data Protection Authority (the APD) in a case we will refer to as the 'IAB Europe decision' (APD, 2022).

The APD handed down a decision in February 2022 as the lead supervisory authority under the 'one-stop-shop' mechanism of Article 56 GDPR. Its judgment was reviewed and approved by a number of Concerned Supervisory Authorities representing the Netherlands, Latvia, Italy, Sweden, Slovenia, Norway, Hungary, Poland, Portugal, Denmark, France, Finland, Greece, Spain, Luxemburg, Czech Republic, Austria, Croatia, Cyprus, Germany (Berlin, Rhineland-Palatinate, North Rhine, Westphalia, Saarland, Lower Saxony, Brandenburg, Mecklenburg-Western Pomerania and Bavaria) and Ireland.

This was not a judgment of the Court of Justice of the European Union, or indeed any other European court. Nonetheless, the breadth of data protection authorities represented – and the consequent scale of the litigation – makes the decision an important precedent within Europe, particularly within the world of online behavioural profiling.

Interactive Advertising Bureau Europe (IAB) is a federation of approximately 5,000 companies across Europe. IAB developed a Transparency and Consent framework as a best practice standard so that real-time bidding could be conducted in compliance with the GDPR (in theory).

Real-time bidding (RTB) was deemed sufficiently complex that it required introduction at the outset of the decision, with diagrammatic representation of the interactions. A distinction was drawn with 'traditional' advertising, in which the advert is negotiated manually between business and publisher. Instead, the machinations of RTB take place 'behind the scenes', with data subjects unaware of the identity of actors involved or even necessarily aware that their information is being automatically auctioned for the opportunity of advertising to them.

The profiling involved in RTB was deemed to be a key element of the processing that IAB had facilitated. There was no controversy that the data used for and generated by this profiling were personal data. This is interesting, as the information used for RTB was very heterogenous, potentially including:

> *URL of the visited site ▪ category or subject of the site ▪ operating system of the device ▪ browser software and version ▪ manufacturer and model of the device ▪ mobile operator ▪ screen dimensions ▪ unique user identification set by vendor and/or buyer. ▪ unique person identifier from the Ad Exchange, often derived from the Ad Exchange's cookie. ▪ the user identification of a DSP, often derived from the Ad Exchange's cookie that is synchronised with a cookie from the DSP's domain. ▪ year of birth ▪ gender ▪ interests ▪ metadata reporting on consent given ▪ geography ▪ longitude and latitude ▪ post code*

While some data included in the RTB processing are what would conventionally be deemed an identifier (gender, post code, year of birth), others are more device-orientated and not 'personal' in the conventional sense (e.g. screen dimensions, browsing software, etc.).

The element of controversy, however, lay in the TC string. The TC string is 'a character string consisting of a combination of letters, numbers and other characters' (para. 41). At paragraph 95 of the judgment, the APD (2022) found that:

> the generation of the TC String in itself constitutes, without any doubt, processing of personal data. The issue at hand is the automated creation, by a CMP registered with the TCF, of a unique and linked set of characters intended to capture a specific user's preferences regarding permitted data exchanges with advertisers. (emphasis added).

The ultimate determination by the APD that the unique set of characters capturing a user's preferences constituted personal data was transformative for the digital economy, acknowledging a whole new link in the chain of information as personal data in and of itself.

The APD's decision is congruent with the logic of this chapter. Although the relevant combination of numbers, letters and characters may not resemble the person in question in a way we would see them with human eyes, in an automated context, this string represents an actionable personal characteristic: their preferences regarding data exchange. It constitutes information that could impact upon the privacy of the person's internet browsing and is therefore, understandably, an identification.

It is important to remember that an identity for the sake of data protection law may be very different from the social, 'real-world' ways we recognise and differentiate people. Identification does not need to include a name or the capacity to physically locate the individual in the real world but could reveal enough information about them to provide an interface to affect

them. McMahon and others illustrate this with the scenario of a woman who miscarries but then continues to receive ads targeted to her perceived pregnancy; a digital profile does not need to correlate accurately with a lived reality to have an impact on her (McMahon et al., 2020). Accurate or not, it would therefore make sense for this profile to be a protected digital identity in order to protect the natural living individual who will be impacted by it.

In this sense, it would not matter if the digital profile correlated poorly with the 'real-world' or 'offline' identity of the individual. Writing for the BBC, Carl Miller conducted a number of subject access requests and uncovered a strange array of inferential judgments made about him based on his browsing history, including that he was a woman trying to conceive, a 'love aspirer' and a disengaged worker with little perceived interest in reading (Miller, 2019). Even if the digital profile of an individual bears little relation to the individual's social or physiological identity, or their own subjective sense of self, it could nonetheless have consequences for them at least in terms of personalised advertisements and (as in the case of misidentification) may have all the more consequences for being wrong. When inaccurate information impacts upon individuals, there is no need to have recourse to the concept of 'fake privacy' (Burgess, 2018) if the digital identity is understood as the clusters of data that can impact a natural, living person.

The IAB Europe case illustrates the increasing penetration of the internet into our daily lives and the consequent expansion of online activity among the digitally connected majority of Europeans, meaning that many of us have an increasing proliferation of 'virtual identities' (Wachter, 2018). Any attempt to rationally delineate those virtual identities that are sufficiently connected with us to constitute a 'profile', and those sufficiently detached to be a 'pseudonym', reveals the lack of attention generally given to the question at the heart of the scope of data protection law: what is an identity in information?

If privacy and data protection are inherently connected to the 'integrity of information constituting one's identity', we cannot understand the boundary of personal data without a common agreement on what information *is* our identity. The general complacency on this issue stems from an apparent assumption that it must be obvious, that 'identity is identity' (Sullivan, 2011). The Spanish AEPD and the European Data Protection Supervisor recently collaborated to address common misunderstandings relating to anonymisation, but the ensuing guidance still falls into the 'identity is identity' trap, stating 'direct identifiers are somewhat trivial to find, indirect identifiers, on the other side, are not always obvious' (AEPD, 2021).

Our exploration of profiling versus pseudonymisation in this chapter shows that direct identifiers are *not* always trivial to define. The evolution of case law since 2016 has shown an expansion of what is considered direct identification in an online environment due to increasing recognition of the power of online profiles – even those that cannot be attributed to the 'real-world' identities of named, gendered, geographically located individuals.

3.2. Pseudonyms as 'Indirect' Identification

It is potentially confusing that a 'pseudonym' can superficially appear the same as a profile, which is also a string of letters and characters. The reason why pseudonymised data are not, however, a direct identification is that they should not permit scrutiny or other action vis-à-vis an individual (e.g. authorising the sharing of their data, in the above example). The French Data Protection Authority (the CNIL) provides the following example:

> an economics researcher has entered into a partnership with a family allowance fund (CAF) which has databases containing the names, dates of birth and addresses of applicants for housing allowance in 2019, as well as the amounts of allowances received and the number of people in the household.
>
> In order to carry out this research and meet data protection requirements, the researcher and CAF have agreed that the latter works on pseudonymised data. For this, the CAF will replace the names and dates of birth with a unique identifier (instead of deleting the columns) and will replace the complete addresses with only the municipalities.
>
> It will thus be possible for the researcher to compare identifiers between databases to find common recipients, without being able to know their identity directly. (CNIL, 2022; emphasis added)

In the above example, the researcher is crucially concerned with *trends across a dataset* rather than scrutinising or making decisions about any individual within it. As such, even if the 'unique identifier' pseudonym was similar in composition to the TC string, its presence within pseudonymised data as opposed to profiling data means that it does not immediately reveal anything about an individual that interferes with their privacy. It is only the risk of 'indirect' identification through combination with other information

that makes this information personal data: it is not an identification in and of itself, as it does not directly impinge on privacy.

3.3. Direct and Indirect Identification

In the above examples, the distinction between 'direct' and 'indirect' identification is key. Direct identification requires no further information and therefore means that the data in question are a legally protected identity without the risk of further attribution. As we have seen above, the French CNIL has referred to pseudonymisation as representing a risk of 'indirect identification', and the UK Parliament has undertaken to go a step further by placing this distinction into law, in proposed updates to its Data Protection Act 2022:

> (3A) An individual is identifiable from information 'directly' if the individual can be identified without the use of additional information.
>
> (3B) An individual is identifiable from information 'indirectly' if the individual can be identified only with the use of additional information. (UK Parliament, 2022, p. 2)

The UK has even gone as far as to propose its own definition of pseudonymisation to clarify that which was set out in the GDPR, indicating that '"pseudonymisation" means the processing of personal data in such a manner that it becomes information relating to a living individual who is only indirectly identifiable' (UK Parliament, 2022, p. 3). While this is only one national interpretation of the GDPR, it does chime with the logic of the CNIL's pseudonymisation scenario, cited above. This helps to reinforce the idea that a pseudonym falls short of a direct identification because it is not immediately revelatory about an individual in a way that will interfere with their rights.

In all EU jurisdictions, the definition of identity will also establish the parameters of data protection law, which protects identified and identifiable people. The scope of this law should be understood with reference to its central purpose: the safeguarding of individual rights within a free market of digital information. Where these rights are engaged by the collection, construction or inference of information, the data should be considered an identification. The difference between pseudonymisation and profiling illustrates this acid test of intrusion in practice.

Data that have undergone GDPR pseudonymisation should not permit evaluation of personal characteristics; they should only reveal trends across individuals. Where reasonable likelihood of attribution back to particular people is removed (though control of the data environment), it may be possible for such pseudonymised personal data to be rendered anonymous. However, careful consideration should be given as to whether the same information could permit profiling in a different context; through combination with other information, or through automated scrutiny with advanced algorithms. These are among the risks of identification that must be excluded by any means reasonably likely to be used for the information to be considered anonymous, per Recital 26 GDPR.

Clarifying the digital identity as distinct from a 'pseudonym' is not just an academic exercise: our privacy and data protection rights are bound up in this concept. We therefore use profiling as a case study of intrusion and impact, which illustrates when information is of such intrinsic value that it constitutes an aspect of identity, thus warranting legal protection.

4. Profiles, Pseudonyms and Anonymity

We have previously written a paper in which we explored the introduction of the 'pseudonymisation' to data protection law within the GDPR. We argued that the data 'environment' (which includes other data, people, the presence or absence of information governance controls and infrastructure) can be managed to render such unattributed information functionally anonymous in the hands of a third party who has no access to the identifiers (Mourby et al., 2018). The controversy surrounding this question continues. Our argument drew on the concept of 'functional anonymisation' and appears to align with the UK Information Commissioner's Office draft updates to their anonymisation guidance post-GDPR (Elliot et al., 2016), but the 'bigger picture' from the European Data Protection Board (EDPB) is still outstanding, as the EU-wide board of regulators is still reviewing the 2014 European guidance on anonymisation (EDPB, 2021).

The preceding sections have shed light on the distinction between profiles and pseudonyms, which forms a central question of this chapter. We can perhaps summarise how this distinction maps onto the personal-anonymous data boundary in Text Box 14.1:

Profiles, Pseudonyms and Anonymity

Profiles: a collection of information with the potential to impact the rights to privacy and data protection of one or more natural persons through automated evaluation of personal characteristics. Profiles thus relate to an 'identified' individual and do not need any further attribution to constitute personal data.

Pseudonyms: information that has undergone GDPR pseudonymisation will still be personal if it can be attributed back to individuals through means reasonably likely to be used (rendering them identifiable per *Breyer*).

Text Box 14.1

To anonymise information, therefore, it is necessary to eliminate:

- Reasonable means of attributing information to individuals through management of the data environment (to prevent the subject becoming *identifiable*).
- The capacity of the information itself to allow individuals to be profiled and thus *identified*.

It is worth noting that longitudinal data that show an individual's behaviour over time (e.g. from a tracking cookie) will be much more difficult (if not impossible) to anonymise than a list of 'hits' on a website. Even if both types of information involve hashed or masked IP addresses, the former is far more likely to enable profiling and therefore remain personal data.

The GDPR could be described as a missed opportunity to provide a clear definition of anonymity versus pseudonymity, and indeed to address the underlying definition of what constitutes 'identification'. As it stands, however, the reader must parse an implicit definition from Recital 26:

> The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an 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.
>
> The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

Elsewhere we have outlined at length how definitions of anonymisation and pseudonymisation can be gleaned from this recital (Mourby et al., 2018). In essence, data that can be attributed to a natural person by means reasonably likely to be used are *indirectly* identifying and are thus pseudonymous personal data. Anonymous data are data for which identification by any means reasonably likely to be used is considered remote. The length of Recital 26 alone illustrates the complexity of demarcating personal and anonymous data in a way that is both logically consistent and consistent with the terminology of the GDPR. This was not unavoidable, however. When reviewing a draft of the GDPR, the Committee on Civil Liberties, Justice and Home Affairs of the European Parliament recommended clarification of these concepts back in October 2012:

> In order to reach the best level of data protection and enable new business models, we need to encourage the pseudonymous and anonymous use of services. Clearly defining 'anonymity' should also help data controllers understand when they are outside the scope of the Regulation. For the use of pseudonymous data, in sense of the data controller is able to single out individual persons by a pseudonym, there could be alleviations with regard to obligations for the data controller. (LIBE, 2012)

To reconcile this paragraph with our working definitions of profiles and pseudonyms, the mere 'singling out' of a person by reference to a pseudonym could be seen as falling a step short of evaluating their personal characteristics in a privacy-intrusive way. As such, it remains logical to see pseudonyms as indirectly identifying personal data, even when they

permit singling out. This appears to have been borne out by the trends we have identified in regarding pseudonymised data as indirect identification.

In short, as pseudonymised data are only personal because of the risk of further attribution, they can be anonymised by eliminating reasonable risk of connection with additional information. Profiling data, however, are directly identifying and cannot be anonymised unless they are modified to the point that they no longer permit the immediate evaluation of personal characteristics.

5. Conclusion

This chapter has suggested that 'identity' in data protection law should be understood not in the psychological sense of how we perceive ourselves but in the 'digital' sense of information with sufficient potential impact on us individually that it should be recognised as a legally protected aspect of self. Although we have focused on profiling as an intrusion into privacy that thus constitutes an identification, the engagement of other fundamental rights could also justify treating the data as personal. For example, where the automated evaluation is of personal characteristics protected under equality laws, identification due to the engagement of the right to non-discrimination should also be considered.

The question of whether information constitutes an identification can thus be considered in two stages:

- Does the information, in and of itself, provide enough detail about the individual that they can be profiled, scrutinised, judged or otherwise experience (even without their knowledge) consequences from this information? If so, they have been 'identified' by the information.
- Can it be combined with other information – either already in the hands of the controller, or which they can obtain through means reasonably likely to be used – in such a way to achieve identification? If so, the individual is 'identifiable'.

Although the GDPR does not explicitly link the definition of profiling with that of personal data, the decisions we have reviewed have placed interference with individual rights at the heart of the concept of identification. As such, profiling provides an important illustration as to when information is sufficiently intrusive into fundamental rights in and of itself that can

justifiably be called an identification. This has been contrasted with pseudonymisation, in which case the question of identification is less certain.

We have therefore considered the theoretical underpinning of the concept of identity in data protection law but also provided some practical guidance. In particular, our analysis highlights that longitudinal data that show individual behaviour over time (e.g. from a cookie) will be much more difficult to anonymise than a logfile of website visitors that only provides a single snapshot in time. Ultimately, however, our central contribution has been to show that it may now be helpful to determine the scope of identity in data protection law with reference to fundamental rights, and not (as is often suggested) the other way around. For all that the category of 'identity' shifts as technology evolves, the underlying benchmarks of privacy and non-discrimination rights are sufficiently stable to provide a reliable sense of who we are as we navigate the digital environment.

References

Agencia Española Protección Data & European Data Protection Supervisor (AEPD). (2021). *10 misunderstandings related to anonymisation*. https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf

Authorité de protection de données (APD). (2022). Litigation Chamber, Case DOS-2019-01377, Concerning: Complaint relating to Transparency & Consent Framework, Decision on the merits 21/2022 of 2 February 2022

Burgess, M. (2018). *The law is nowhere near ready for the rise of AI-generated fake porn*. Wired. https://www.wired.co.uk/article/deepfake-app-ai-porn-fake-reddit

CNIL. (2022). *Scientific research* (*excluding health*)*: Challenges and advantages of anonymization and pseudonymization*. https://www.cnil.fr/fr/recherche-scientifique-hors-sante/enjeux-avantages-anonymisation-pseudonymisation

Committee on Civil Liberties, Justice and Home Affairs of the European Parliament (LIBE). (2012). Working Document 2 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). https://www.europarl.europa.eu/doceo/document/LIBE-DT-497802_EN.pdf?redirect.

de Hert, P., & Lammerant, H. (2016). Predictive profiling and its legal limits: Effectiveness gone forever? In B. van der Sloot, D. Broeders, & E. Schrijvers (Eds.), *Exploring the boundaries of Big Data* (pp. 145–167). Amsterdam University Press.

Elliot, M., Mackey, E., & O'Hara, K. (2016). *The Anonymisation Decision-Making Framework*, 2nd ed. UKAN Publications. https://ukanon.net/framework/

European Data Protection Board (EDPB). (2021). *EDPB letter to the European institutions on the privacy and data protection aspects of a possible digital euro*. https://edpb.europa.eu/system/files/2021-07/edpb_letter_out_2021_0112-digitaleuro-toep_en.pdf

Information Commissioner's Office. (2020). *What is personal data?* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/

Information Commissioner's Office. (2022). *Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance*. https://ico.org.uk/media/about-the-ico/consultations/4019579/chapter-3-anonymisation-guidance.pdf

Laurie, G. (2002). *Genetic privacy: A challenge to medico-legal norms*. Cambridge University Press.

McMahon, A., Buyx, A., & Prainsack, B. (2020). Big data governance needs more collective responsibility: The role of harm mitigation in the governance of data use in medicine and beyond. *Medical Law Review*, *28*(1), 155–182.

Miller, C. (2019). *Would you recognise yourself from your data?* BBC. https://www.bbc.co.uk/news/technology-48434175

Mourby, M., et al. (2018). Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law and Security Review*, *34*(2), 222–233.

Oxford University Press. (2007). *OED*. https://www.oed.com/

R. (on the application of Bridges) v Chief Constable of South Wales. (2020). EWCA Civ 1058.

Regulation 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/ (General Data Protection Regulation). (2016). OJ L119/1.

Sullivan, C. (2011). Digital identity: An emergent legal concept. University of Adelaide Press.

UK Parliament. (2022). Data Protection and Digital Information Bill. https://publications.parliament.uk/pa/bills/cbill/58-03/0143/220143.pdf

Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law and Security Review*, *34*(3), 436–449.

Part 2 Synopsis: Article 8(1) and Private Life

2.1 What is 'Private Life'?

A 'privacy-based approach to private life' may sound tautologous, but as 'private life' is the next key part of Article 8 ECHR to consider, Part 2 of this thesis now seeks to ground this concept within the potential for interference with patients' rights to private and family life.

As previously explained, this thesis follows the structure of Article 8. Part 1 explored the initial, threshold question: is an individual identified by information? I concluded that the boundaries of 'identity' in information should be determined according to whether its use can interfere with an individual's private life. Private life was understood expansively, to include the potential for an individual to be scrutinised, profiled and individually evaluated, even if their 'real-world' self is not revealed in the process. If this threshold requirement is made out, I argued that the information in question should be considered personal data, and data protection law applied to the use of the relevant information.

This question of identification is only the first step, however, for other areas of information law. As this Part 2 explores, under both the 'common law' of confidentiality in England, and the emerging MOPI tort, information must be not only personal data, but also 'private' to engage Article 8 ECHR. Part 2 of this thesis, therefore, further develops my account of what 'private life' should mean, in the context of secondary uses of patients' data. This term comes directly from the text of Article 8:

> Everyone has the right to respect for his private and family life, his home and his correspondence.

There is no succinct definition as to what is meant, here, by 'private and family life.' In *S and Marper v the United Kingdom*,[148] a case concerning government retention of genetic data, the European Court of Human Rights ('ECtHR') observed:

> the concept of "private life" is a broad term not susceptible to exhaustive definition. It covers the physical and psychological integrity of a person [..]It can therefore embrace multiple aspects of the person's physical and social identity […] Information about the person's health is an important element of private life[149]

There is some overlap between this broad characterisation of private life, and the concept of personal data. The Council of Europe has also set data protection principles, which form another bridge between the EU concept of personal data and the ECHR concept of private

[148] Introduction, note 4.
[149] Ibid at [66].

life.[150] But there is still some ambiguity as to how private information may differ, conceptually, from personal data. As the ECtHR has not 'exhaustively' defined the concept of private life, the English courts have developed their own gloss as they have applied Article 8 ECHR to private and confidential information.

Indeed, as I will show, not all information which qualifies as personal data has been deemed private by the English courts. English case law has determined that, to qualify as private or confidential, information must not only identify people, but also attract a 'reasonable expectation of privacy.'[151] This test has the potential to lead scrutiny *away* from the proper scope of private life, and into questions of how logical or rational a claimant may have been in their expectations of privacy. Consequently, legal protection can become limited to 'reasonable' patients, rather than to all patients whose private lives may be affected by the use of their information. As I explain in subsection 2.2 below, this then leaves patients at the mercy of judicial constructions of the broad concept of 'reasonableness.'

To assess whether it is helpful to apply a 'privacy-based approach' to the scope of private life, I explore a different way of determining whether patients' information should be considered private. Applying the working definition of 'privacy' adopted in my thesis, I will ground the concept of Private Life in the potential for interference with Article 8 ECHR. I attempt this through exploration of two case studies, within one publication.

2.2 Publication 4: Reasonable Expectations Article

Publication Title: Private By Default: Reasonable Expectations in Secondary Uses of Patient Data

Publication Type: Journal Article

Authorship: Sole

This sole-authored article was submitted to the *Medical Law Review* in March 2025. In May 2025, the peer-reviewer recommended publication, subject to major revisions. At the time of writing (July 2025), I have made the recommended revisions, and plan to resubmit it in August 2025. The '4.reasonable expectations article' argues that, by default, all personal data the NHS holds about its patients should be considered private and confidential. In particular, it rebuts the idea that such information needs to be 'health-related,' and suggests instead that it should be presumptively reasonable to expect privacy in all your identifiable NHS data.

---

[150] Council of Europe Convention of 1981 for the protection of individuals with regard to automatic processing of personal data.

[151] For the MOPI tort, this case been confirmed by the *Supreme Court in Bloomberg LP v ZXC*, [2022] UKSC 5; [2022] AC 1158, [42]. For the 'common law' duty of confidence, the position is less clear, but this test was applied by the Court of Appeal [2015] EWCA Civ 1034 (Introduction, note 88). This is explored in detail in the 4.reasonable expectations article.

This argument is made in the context of two case studies: the **Immigration Case Study** and the **DeepMind Case Study**.

A. The Immigration Case Study

This example centres around the Court of Appeal's judgment in in *R (on the application of W, X, Y and Z) v Secretary of State for Health and Secretary of State for the Home Department.*[152] This litigation emerged from a seven-year period (2011-2018), during which the NHS shared patients' data with the Home Office for the purpose of enforcing immigration sanctions. Following a change in immigration rules in 2011, non-citizen NHS patients who owed money to the NHS (from unpaid charges for treatment) could be prevented from re-entering the UK.

The information provided from the NHS to the Home Office was not necessarily 'health data,' in the sense that it did not reveal the detail of these patients' medical treatment.[153] But, as I show in this article, the information clearly had the capacity to impact the private and family lives of the affected patients. Such information, in the hands of the Home Office, would enable the latter to make decisions about them with serious legal consequences (affecting, for example, their capacity to re-enter the UK).[154] These decisions may have significant implications for private life: in particular, for the patients who may have been prevented from re-joining their friends and family in the UK, and thus 'establish and develop relationships with other human beings.'[155]

As I argue in the 4.reasonable expectations article, this does not mean that the disclosure of information to the Home Office could not have been justified under Article 8(2) (which is considered further in Part 3 of this thesis). In this '4.reasonable expectations article,' though, I focus on the scope of 'private' information under Article 8(1) ECHR. The information presented to the High Court, and the Court of Appeal, in this litigation indicated ample grounds that the patients' private lives could have been impacted by the disclosure to the Home Office. The courts' conclusion that these data were not, in fact, private and confidential because the claimants lacked a reasonable expectation of privacy is concerning and, I suggest, out of step with the proper scope of Article 8. I argue that the Court was misguided in its insistence on the 'reasonable expectations' test, when there were far more compelling grounds on which to determine the application of Article 8. The potential for impact on private and family life is at the core of the interests the right is designed to protect, and the transmission of information to the Home Office had a non-trivial chance of impacting the claimants' ability to re-enter the UK and see their families. I therefore argue for a

[152] [2015] EWCA Civ 1034, note 88.
[153] Although one of the claimants ('Z') accidentally had the details of her treatment disclosed, which the High Court held to be 'obviously unlawful', see *R (on the application of W, X ,Y & Z) v Secretary of State for Health* [2014] EWHC 1532 (admin); (2014) WL 1220234, at [61].
[154] Under Article 22 GDPR, for example, any decisions with legal consequences about a data subject would qualify as 'significant.'
[155] Another aspect of the right to private life emphasised in *S and Marper v UK* (Introduction, note 4) [66].

rebuttable legal presumption of reasonable expectations of privacy in secondary uses of patients' identifiable data.

B. The DeepMind Case Study

The second case study in the 4.reasonable expectations article revolves around a large-scale programme of secondary uses of patients' data, which culminated in the Court of Appeal's judgment in *Prismall v Google UK Ltd.*[156]

In 2015-16, the Royal Free NHS Trust disclosed the data of 1.6 million patients to DeepMind Technologies Ltd (or 'Google DeepMind,' as it was sometimes called in acknowledgement of its parent company). The stated purpose of sharing this information was the development of an app to identify patients with acute kidney injury ('AKI'). Of the patients who had their information disclosed, only a very small proportion would have had AKI, or any associated symptoms. As Hal Hodson and Julia Powles noted in their review of this disclosure, a large proportion of the data shared would thus not have been relevant to the diagnosis of AKI, raising concerns about the proportionality of the interference with patients' privacy.[157]

There was no dispute that the patients whose records were disclosed were identifiable, and that Google DeepMind had therefore received personal data. Indeed, the Information Commissioner's Office was very critical of the disclosure from a data protection perspective.[158] However, the representative patient claimant, Andrew Prismall, brought a case under the 'MOPI' tort, and not under the GDPR.[159] The claim was ultimately rejected because the hypothetical 'lowest common denominator patient' among them lacked a reasonable expectation of privacy.

This case study highlights the distinction between 'personal data' and 'private information' within English law that I am critiquing in this Part 2 of my thesis. Collapsing the sharp distinction between the two concepts, and emphasising their common source in privacy, would help avoid the anomalous conclusion that these 1.6 million patients did not have Article 8 rights in their data. Had the 'reasonable expectations' test been de-centred— as my privacy-based approach requires— these judgments might have aligned better with the Information Commissioner's conclusions.

2.3 Contribution of Part 2

Part 2 of this thesis argues that all personal data the NHS holds about its patients should be considered 'private information' by default. As personal data is (per my definition from Part

---

[156] [2024] EWCA Civ 1516, Introduction note 44.
[157] J. Powles and H. Hodson, 'Google DeepMind and healthcare in an age of algorithms' (2017) 7 Health Technology, 351-367.
[158] E. Denham, Introduction note 46.
[159] Possibly due to the difficulties experienced by the claimants attempting to bring a class-action under the Data Protection Act 1998 in *Lloyd v Google* (see below, note 160), but this is my speculation.

1) information which interferes with an individual's privacy, the subsequent question of 'reasonable expectations' of privacy becomes less significant. This argument is made in the context of secondary uses of health data, as that is the scope of this thesis. However, it could have broader implications for the relationship between personal data and private information— and therefore for the relationship between data protection law and the MOPI tort.

Part 2 ultimately provides an affirmative response to my research question: 'can a privacy-based approach help regulate secondary uses of patients' data?' My 'privacy-based' approach can provide a more robust understanding of patients' privacy interest in their personal information. Doctrinally, it creates better coherence between the concepts of personal data and private information, which share the common source of Article 8 ECHR.[160] My approach could also help close loopholes in patients' privacy, currently created by the 'reasonable expectations' test, as illustrated by the two case studies discussed above.

The overall contribution of Part 2 of this thesis, therefore, is the argument that Article 8 ECHR should apply, by default, to secondary uses of patients' identifiable data. The implications for how these secondary uses should therefore be justified, by default, under Article 8(2) will be considered further in Part 3.

[160] See, for example, *Lloyd v Google LLC* [2021] UKSC 50; [2022] A.C. 1217 at [111]— although a difference in approach between the data protection and common law privacy regimes will be acceptable where the former has clear statutory rules which deviate from the courts' general approach under Article 8 (as here, on the issue of damages).

Publication 4

Publication Title: Private by Default: Reasonable Expectations in Secondary Uses of Health Data

Short Title: ‘Reasonable Expectations Article’

Page Number of Thesis: 63

The following pages incorporate a PDF of a work-in-progress version of this article. This version has been accepted for publication in the *Medical Law Review* but the article has not yet been published on an open-access basis. A request to cover the fees for open-access publishing is currently being processed. I have removed this article from the public version of this thesis, for now.

The formatting, pagination and footnote numbering of the publication have been retained, and thus the numbering is self-contained. The pagination sequence of this thesis will then resume at the end of the publication.

Part 3 Synopsis: Article 8(2) and Justification

The word 'justification', as used in this thesis, is essentially a shorthand for the requirements of Article 8(2) ECHR, which stipulates:

> There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society [...].

Therefore, even if secondary uses of patients' information are an 'interference' with their rights under Article 8 ECHR, this interference can be justified if it is in accordance with the law and 'necessary in a democratic society.' Part 3 of this thesis explores the implications of using Article 8(2) as a benchmark to evaluate and justify secondary uses of patients' data.

The final third of my thesis is, thus, no longer concerned with the *scope* of patients' protections under information law, but rather the *nature* of this protection. In particular, the nature of justification under Article 8 ECHR when it is engaged by the secondary use of patients' data. The research question of Part 3 can thus be summarised as: '*what difference does the application of Article 8 make to the way secondary uses of patients' data should be justified?*'

To answer this question, Part 3 will focus mainly on the GDPR, as containing the most detailed regulatory framework governing how to evaluate and justify secondary uses of personal data. It is in the interpretation of the GDPR's requirements, as they apply to patients' data, that I suggest the application of Article 8(2) can make a more concrete contribution (see 3.1, below).

This Part is made up of three publications: two journal articles and a book chapter. I will explain the contribution of each publication individually, but they can be summarised as focusing on the justificatory requirements of:

1) Proportionality (publications 5 and 6), and
2) Non-discrimination (publication 7).

The 5.academic governance article and the 6.EHDS chapter establish that Article 8 ECHR imposes its own requirements of necessity and proportionality on secondary uses of patients' data. These requirements mirror and the GDPR's own obligations of necessity[161] and proportionality[162] which might otherwise be limited by scientific research exemptions (as considered within the **Scientific Research Case Study**).

[161] Particularly the requirement under Article 6 that processing be necessary for a specified lawful purpose.
[162] See Introduction note 51 for a definition of the term 'proportionality' in this thesis, i.e. 'whether an interference with (inter alia) Article 8 ECHR is necessary in a democratic society, and represents no greater

Article 8 ECHR is also supplemented by the anti-discrimination principles of Article 14 ECHR. The requirement not to discriminate on legally protected grounds is an ancillary obligation to the right to private life. The 7.DPIA+ article argues that combined consideration of Articles 8 and 14 opens data protection law (which is often seen as individualistic) to considering groups of patients with the same legally protected characteristics. This is explored within the **Covid-19 Case Study.**

The conclusion of this Part 3 is that the proportionality and anti-discrimination principles derived from Article 8 ECHR do indeed help fortify the GDPR's justificatory requirements for secondary uses of patients' data. Therefore, the application of Article 8(2), as a final stage of the 'privacy-based approach' of this thesis, can also help to regulate secondary uses of patients' data.

3.1 Justification under MOPI, Confidentiality & Data Protection

The focus of the three publications in Part 3 is how secondary uses of patients' data[163] should be justified under the GDPR, bearing in mind the additional requirements of Article 8(2) ECHR. As explained above, this is because the GDPR has more detailed provisions governing the justification of patients' data, compared to the common law of confidentiality and MOPI. As such, a concrete contribution on justification can feasibly be made here by using Article 8(2) as a benchmark for interpreting the GDPR's provisions.

Outside of my thesis publications, I have also considered how the engagement of Article 8(2) could strengthen the public interest justifications available under the MOPI tort, and the common law of confidentiality. However, this common law justification is construed so broadly by the courts that it was difficult to make as concrete a contribution as I could by focusing on the GDPR. The courts have broad discretion in how they can construct a public interest justification under a confidentiality or MOPI claim, and can make a finding of justification with minimal analysis. Within the *Prismall* class action of the **DeepMind Case Study,** for example, the High Court dismissed the justificatory requirements of Article 8, by suggesting (incorrectly, in my view) that they did not require any additional analysis to the question of 'reasonable expectation of privacy':

> The scenario and factors that fall to be considered when assessing whether the interference was justified are those I have identified when addressing the prospects of all in the Claimant Class showing a reasonable expectation of privacy, including the

intrusion into a Convention right than is necessary for the legitimate aim pursued. There are multiple requirements within the GDPR that mirror the ECHR's broad model of proportionality as no more than necessary for a particular purpose. Key among these would be the data minimisation principle, and the purpose specification principle, in Article 5. These are discussed in detail in the 5.academic governance article (**thesis page 75**) and the 6.EHDS chapter **(thesis page 76**).

[163] Or, in the case of the 5.academic governance article, patients' data as a category of research data used by university researchers.

> extent to which the alleged interferences did or did not come within direct care. It is unnecessary to repeat that analysis.[164]

I explained in Part 2 why the Court's approach to 'reasonable expectations' in this case was insufficient to capture the proper scope of Article 8(1). I would also suggest that the 'reasonable expectation' factors considered there are insufficient to cover the requirements of Article 8(2) for interference to be necessary and proportionate (see below, 3.2).

A perfunctory approach to Article 8(2) is not unique to the litigation in the **DeepMind Case Study**, however. Within the **Immigration Case Study**, the Court of Appeal made a finding of justification under Article 8(2) on the basis that the relevant data-sharing powers were (at a broad, national level) in accordance with UK law.[165] The impact on the claimants' private and family lives, and the proportionality of this impact to the policy aim pursued, were not factored into the Court's analysis in this judgment.

What the Court of Appeal *did* take into account, in their finding on Article 8(2), was the fact that the data-sharing powers in question were subject to the (then) Data Protection Act 1998.[166] The Court suggested that the application of data protection law was a key safeguard to prevent the unlawful or disproportionate use of patients' personal data. The judges did not engage in any detailed analysis as to whether the processing of the claimants' information had (on the facts of this case) complied with data protection law. The mere existence of these statutory safeguards was taken as enough to render the disclosure 'in accordance with the law' for the sake of Article 8(2) ECHR.

If data protection law is taken to be a broad safeguard that justifies secondary uses of patients' data under Article 8(2) ECHR, its requirements should be considered in more detail. I suggest that the mere existence of the GDPR does not mean that all secondary uses of patients' data will be necessary and proportionate. Rather, the GDPR is only an adequate safeguard for patients' rights if it is interpreted and applied with sufficient regard for proportionality and anti-discrimination requirements.

As I explain in my publications, concerns have been raised about the implementation of the GDPR within (what I term) my **Scientific Research Case Study**. These questions include whether the GDPR's research exemptions go too far,[167] create too much fragmentation along national lines,[168] and, more broadly, whether data protection systems do not capture wider

---

[164] *Andrew Prismall v Google UK* [2023] EWHC 1169 (KB), [172].
[165] [2015] EWCA Civ 1034, note 88, [75]-[87].
[166] Ibid, [87].
[167] K. Pormeister, 'Genetic data and the research exemption: is the GDPR going too far?' (2017) 7 International Data Privacy Law 2, 137-146. This is discussed further within the 5.academic governance article (**thesis page 75)** internal page 25.
[168] European Commission, Directorate-General for Health and Food Safety, 'Assessment of the EU Member States' rules on health data in the light of the GDPR' (12 February 2021) https://health.ec.europa.eu/system/files/2021-02/ms_rules_health-data_en_0.pdf. This is discussed in the 6.EHDS chapter **(thesis page 76)** internal page number 5.

risks to society arising from data processing.[169] The conclusion I draw over the course of three publications, in Part 3, is that the application of Article 8(2) to our interpretation of the GDPR can help mitigate these criticisms of data protection law.

3.2 Publication 5: The Academic Governance Article

Publication Title: Governance of Academic Research Data Under the GDPR— Lessons from the UK

Publication Type: Journal Article

Authorship: Miranda Mourby, Heather Gowans, Stergios Aidinlis, Hannah Smith, Jane Kaye

Background

This '5.academic governance article' was published in *International Data Privacy Law* in 2019. Like the 1.pseudonymisation article,[3] it was written with colleagues on the ESRC-funded Administrative Data Research Network ('ADRN'). As part of this collaboration, we wrote co-authored pieces to reflect our discussions and findings within the project. I had the initial idea for the article, wrote most of the text, and edited it all into a final version. As always, all co-authors read and commented on the article at each stage of its preparation. Stergios Aidinlis and Hannah Smith, who were DPhil students at the time, made significant contributions to the drafts of this article, and their input deserves particular acknowledgement. The core argument for the sake of this thesis, however— that Article 8 ECHR can help mitigate any lacunae created by the GDPR's research exemptions and academic derogations— was my own contribution.

Contribution

The key contribution of this '5.academic governance article' for my thesis is its argument that the application of Article 8 ECHR to patients' data can mitigate any limitations or ambiguities in the GDPR stemming from its scientific research exemptions.

The term 'academic governance' in the title may appear to go beyond the scope of this thesis, which focuses on secondary uses of patients' data. However, as with the 1.pseudonymisation article,[4] the term is broad enough to encompass the secondary use of patients' data by academic researchers, which falls within the **Scientific Research Case Study** of this thesis.

By 'academic data,' we meant personal data processed by researchers employed by universities— particularly for the purposes of scientific research. For example, in the text box in section II,[170] I outline a scenario in which university researchers may need to consider the scope of their derogations from the GDPR. In the article, it is described only as 'sensitive

[169] See Introduction, notes 77 and 80.
[170] 5.academic governance article (**thesis page 75)** internal page number 20.

research personal data,' but this hypothetical scenario was drawn from my experience working as a researcher on medical consortia.[171] Thus, while the article frames its discussion in broader terms than the secondary uses of patients' data, its findings are all relevant to the research question of this thesis, which focuses on secondary use of patients' data. The article explores how the application of Article 8 ECHR to the processing of personal data by public authorities[172] can expand and solidify the protections afforded by the GDPR, particularly where the latter are still novel or uncertain.

As an example, the first text box in section II[173] outlines a hypothetical scenario in which university researchers process a large amount of special category personal data for a scientific project. Under the GDPR and the Data Protection Act 2018,[174] some exemptions would be available if the researchers wanted to modify their GDPR obligations. In the case of their transparency obligations under Article 14, for example, there are some exceptions to the duty to inform data subjects that they have received their data from a third party.[175] The personal data can also be retained for longer periods of time if they are used for scientific research,[176] and will be deemed an automatically compatible purpose with the initial purpose of collection.[177]

The full detail of the academic and scientific exceptions under the GDPR, and the Data Protection Act 2018, is explained further within the 5.academic governance article. For the purposes of the wider thesis, however, the salient point is that under Article 8 ECHR the large-scale processing of special categories of personal data[178] requires robust safeguards to prevent harm such as discrimination.[179] This may not, in fact, require any more safeguards for scientific research than the GDPR already contains.[180] Rather, it requires a more cautious interpretation of the GDPR's scientific exemptions and academic derogations, when they are applied in practice.

The 5.academic governance article thus concludes that privacy rights under Article 8 ECHR raise the bar for the academic and scientific exemptions to data protection law. The puzzle of intersecting law is undeniably complex, but ultimately Article 8 provides a touchstone to

---

[171] The published version of the article acknowledges the support of the Innovative Medicines Initiative Joint Undertaking under grant agreement no115317 ('DIRECT); a project using data collected in collaboration with NHS Trusts to study diabetes.
[172] In this article, the public authorities in question are universities, but the principles are broadly the same as for the NHS itself.
[173] See note 170.
[174] The piece of legislation which sets out the UK's national derogations from the GDPR, including for academic purposes and scientific research.
[175] GDPR, Article 14.5(b).
[176] GDPR, Article 5.1(e).
[177] GDPR, Article 5.1(b).
[178] Which, under Article 9 GDPR, includes information relating to health.
[179] GDPR, Recital 71.
[180] Under Article 89.

understand the scope of patients' rights, even when exceptions to GDPR obligations might otherwise render them ambiguous.

3.3 Publication 6: The EHDS Chapter

Publication Title: Secondary Uses of Patients' Data in the European Health Data Space: A UK-German Comparison

Publication Type: Book Chapter

Authorship: Miranda Mourby, Fruzsina Molnár-Gábor

Background

This '6.EHDS chapter' was prepared for Edward S. Dove's edited book, *Confidentiality, Privacy, and Data Protection in Biomedicine: International Concepts and Issues*, which was published in September 2024.

It was written with Fruzsina Molnár-Gábor, who was the Principal Investigator of the 'TrustDNA' project, on which I worked as a part-time research fellow in 2023-2024. I wrote almost all the text, with Professor Molnár-Gábor reviewing and commenting on several drafts, in particular to add insights into German law. As editor of the book, Professor Dove also reviewed and commented on a draft before accepting it for publication.

This chapter offers a final perspective on the application of Article 8 ECHR to secondary uses of patients' data within the **Scientific Research Case Study.** While previous publications focused on scientific research within the UK, the 6.EHDS chapter takes into account the potential for national variation in the law governing secondary uses of patients' data, and its implications for international research projects. We therefore considered whether legal and constitutional differences between Germany and the UK would have any impact on the way patients' data can be used for secondary purposes within the European Health Data Space (or 'EHDS').

Contribution

The core contribution of the 6.EHDS chapter is its conclusion that the UK and Germany have taken compatible approaches to Article 8 ECHR, at least as regards secondary uses of patients' data. This helps to mitigate the concerns— raised in academic[181] and policy[182]

[181] J. Dumortier and B. Mahault, 'European-Wide Big Health Data Analytics under the GDPR.' In Tzanou (ed) *Health Data Privacy under the GDPR: Big Data Challenges and Regulatory Responses* (Oxford: Routledge, 2021). 56–70; Rossana Ducato, 'Data protection, scientific research, and the role of information' (2020) 37 Computer Law & Security Review 105412; R. Becker, A. Thorogood, J. Ordish, M. J.S. Beauvais, 'COVID-19 Research: Navigating the European General Data Protection Regulation' (2020) 22 Journal of Medical Internet Research 8, e19799.

[182] European Commission, note 168.

literature— that interpretations of the GDPR's scientific research provisions will create national fragmentation, and impede international research using health data. At least in this study, however, our findings were encouraging. We were able to reconcile the two jurisdictions' interpretation of the GDPR *and* Article 8 ECHR, at least as they applied to secondary uses of patients' data. This suggests that my 'privacy-based' approach need not be nationally divisive.

In both countries, an 'opt-out' system— where patients' information will be used for secondary purposes in the absence of any active objection— has been accepted as an appropriate safeguard for patients' individual autonomy. This represents a development from research I had conducted in a previous project, which suggested that Germany might continue to prioritise explicit, 'opt-in' consent for research.[183]

At the outset of writing this chapter, it seemed possible that Germany would take a different approach to secondary uses of patients' data within the EHDS. However, our analysis of the EHDS Regulation suggests this is not necessarily the case. Despite Germany's constitutional differences from the UK, it has accepted the text of the Regulation on the EHDS. This means Germany will also share patients' data for secondary uses on an 'opt-out' basis— that is, without patients providing active consent to the transmission of their information. As the EHDS is constructed on an 'opt-out' basis across the EU, the data of up to 450 million patients stands to be linked for secondary uses without their active consent, or control.

This suggests that the 'privacy-based approach' explored in this thesis may have value outside the UK, to other EU Member States and ECHR-signatory countries. Individuals' choices will not, by themselves, be sufficient to regulate the EHDS, because huge amounts of patient data will be shared without their input. As such, the approach to justification explored in this Part 3, which centres public authority obligations rather than individuals' rights, is poised to have relevance for the forthcoming EHDS.

This means that the 'privacy-based' approach taken in this thesis also has utility for international research projects, where participants' personal data crosses national boundaries. If Article 8 can be understood compatibly in the UK and Germany, this mitigates concerns expressed about the fragmentation of the GDPR as it has been applied nationally,[184] and the issues this has created for international scientific research.

---

[183] K. Ó Cathaoir, E. Gefenas, M. Hartlev, M. Mourby and V. Lukaseviciene, 'EU-STANDS4PM report: Legal and ethical review of in silico modelling' (March 2020) https://www.eu-stands4pm.eu/lw_resource/datapool/systemfiles/elements/files/AA77832F664661DBE0537E695E8689E3/current/document/WP3_March2020_D3-1_V1_public.pdf.

[184] See European Commission (note 168).

3.4 Publication 7: The 'DPIA+' Article

Publication Title: The 'DPIA+': aligning data protection with equality law

Publication Type: Journal Article

Authorship: Sole

Background

This sole-authored article (which I will call the '7.DPIA+ article') was submitted to the *Computer Law & Security Review* in March 2025. At the time of writing, it is under review. It explores the contribution of equality law to data protection tools, focusing on the Data Protection Impact Assessment.

The DPIA+ article explores the added value of equality law for the GDPR's requirement to conduct a Data Protection Impact Assessment or 'DPIA' when processing large amounts of special category personal data (such as health-related data).

Contribution

The key contribution of the 7.DPIA+ article is its argument that evaluative and justificatory exercises under the GDPR can be reinforced by the equality requirements of Article 14, as an ancillary consideration of Article 8(2). This is considered in the context of the **Covid-19 Case Study**. Proportionality and public interest exemptions (which were the focus of the 5.academic governance article and the 6.EHDS chapter) have already been examined in the context of state responses to Covid-19.[185] In order to make an original contribution here, I instead focus on the ECHR's non-discrimination requirements, as an under-explored complement to the GDPR.

The 7.DPIA+ article thus explores the contribution of equality law to the considerations required by a 'DPIA.'[186] I describe my proposed combination of the two as a 'DPIA+.' I focus on equality law as ultimately stemming from Article 14 ECHR, which prohibits discrimination in respect of a convention right.[187] Where Article 8 ECHR applies in relation to the processing of personal data, therefore, Article 14 will also import non-discrimination obligations, which I suggest can enhance the bare bones of the DPIA under the GDPR.

---

[185] See Becker et al, note 181. See also A. Spadaro, 'Covid-19: Testing the Limits of Human Rights' 11 European Journal of Risk Regulation 2, 317-325, and Ada Lovelace Institute, 'Exit through the App Store?' (20 April 2020), available from: https://www.adalovelaceinstitute.org/our-work/covid-19/covid-19-exit-through-the-app-store/ (cited by J. Pila in 'Covid-19 and Contact Tracing: A Study in Regulation by Technology' (2020) 11 European Journal of Law and Technology 2).

[186] Data Protection Impact Assessment, under Article 35 GDPR.

[187] C. McCrudden and S. Prechal, 'The Concepts of Equality and Non-Discrimination in Europe: A practical approach' (2009) https://ec.europa.eu/social/BlobServlet?docId=4553.

The 7.DPIA+ article considers another dimension to the large-scale processing of special category personal data. Where information falling within these categories (such as health-related information) is processed at scale, a data protection impact assessment ('DPIA') should be undertaken. I explore how the requirements of a 'DPIA' under the GDPR can be expanded by combined consideration with equality law. In particular, I consider the public sector equality duty ('PSED') under s.149 Equality Act 2010. While this is only one aspect of equality law within the UK, it has the most natural alignment with the 'DPIA' requirement of the GDPR, as it currently forms the basis for 'Equality Impact Assessments.'[188]

An additional benefit to my 'privacy-based approach' to secondary uses of patients' data, therefore, is that Article 14 ECHR also imposes non-discrimination requirements in respect of patients' rights under Article 8 ECHR. I use the PSED as my point of focus within equality law, to permit sufficiently detailed analysis of a single **Covid-19 Case Study**, and draw upon existing caselaw on the PSED.[189] However, the combined effect of Articles 8 & 14 ECHR would also apply to NHS bodies more directly, through their obligation to comply with ECHR rights under s.6 Human Rights Act 1998.

I describe the net effect of combining a DPIA evaluation with consideration of equality law as a 'DPIA+.' I review its potential benefits with detailed reference to NHS England's Covid-19 data store, which was established in 2020 to help the government plan its response to the first wave of the pandemic. I suggest that it is not necessarily fair or accurate to characterise data protection law, or DPIA's, as concerned exclusively with the interests of individual data subjects.[190] At the same time, however, the combined consideration with the PSED imports more structured evaluation of the impacts of data processing on groups with particular protected characteristics— in this case: age, race and disability.

3.5 Conclusion on Part 3

Part 3 of this thesis explores a variety of benefits to the application of Article 8 ECHR to patients' data. These benefits can be summarised as falling into two main categories:

1) Article 8 ECHR imposes its own requirements of necessity and proportionality to the justification of secondary uses of patients' data, which can fortify those of the GDPR which may be unclear, or narrow, in places;
2) Article 8 ECHR also imports the anti-discrimination principles of Article 14 ECHR, which opens data protection law (which is often seen as individualistic) to considering groups of patients with legally protected characteristics.

---

[188] A non-mandatory form of report in which compliance with the public sector equality duty can be demonstrated via a written record of internal evaluation. See D. Pyper, 'The Public Sector Equality Duty and Equality Impact Assessments' (2020) https://researchbriefings.files.parliament.uk/documents/SN06591/SN06591.pdf.

[189] Particularly *R (Bridges) v The Chief Constable of South Wales & Ors* [2020] EWCA Civ 1058.

[190] See also the characterisation of 'privacy' in this thesis as concerned with values beyond individual autonomy, see thesis Introduction, VI. C, 'Beyond Autonomy.'

These two overarching benefits are the final part of my answer to the question 'can a privacy-based approach help regulate secondary uses of patients' data?' This Part illustrates how the regulation of patients' data can be enhanced by the application of Article 8 ECHR, which provides a foundation for a more expansive understanding of data protection.

The application of Article 8 ECHR, by default, to the secondary uses of patients' personal data leads to a more expansive understanding of 'justification' under data protection law. Understood in isolation, data protection legislation has significant potential gaps. Within the **Scientific Research Case Study**, publications 5 and 7 explore the GDPR's broad exemptions for scientific research, which can modify default requirements such as data minimisation.[191] Within the **Covid-19 Case Study**, publication 6 explores how the GDPR's DPIA requirement[192] risks becoming a toothless bureaucratic exercise.

These potential lacunae in the GDPR's safeguards are of less concern, however, if they are (by default) considered alongside the justificatory requirements in Article 8(2) ECHR. As emphasised in the '5.academic governance article,' the principle of proportionality under Article 8(2) reinforces that of data minimisation under the GDPR. Where it is ambiguous how much the latter applies, because a data controller can claim scientific research exemptions, the principle of proportionality should nonetheless ensure that patients' identifiable data are not used in too great a volume, for too long, or for more disparate purposes than is necessary for the research project in question. Likewise, even if a data controller were minded to pursue only a superficial DPIA, this is unlikely to be sufficient to encompass equality and human rights considerations, which could be consolidated into a broader 'DPIA+.'

As the 7.DPIA+ article emphasises, the GDPR is not just a set of information security requirements, or a menu of data subject rights. It is also an expression of the rights to privacy and data protection, meaning that the use of personal data by public authorities must be necessary and proportionate. As such, a purposive understanding should be taken of any potential 'loopholes' (such as scientific research exemptions), to ensure that they are understood in terms of strict necessity, and ensuring that the least intrusive means of accomplishing the research are pursued.

The anti-discrimination requirements of Article 14 ECHR are another dimension of Article 8 which can complement the bare text of the GDPR. Although Article 35 GDPR expresses minimum questions a DPIA should consider, these could be combined with consideration of

---

[191] Which, echoing the ECHR's proportionality requirements, means that personal data should be 'limited to what is necessary in relation to the purposes for which they are processed.' GDPR, Article 5.1(c). This principle is emphasised as a general obligation in Article 25 (data protection by design and default) and more specifically in the context of scientific research in Article 89, and Recital 156.

[192] Article 35. An assessment 'of the impact of the envisaged processing operations on the protection of personal data,' where processing is likely to result in a risk to people's rights and freedoms. This is explored in more detail in the 7.DPIA+ article.

the impact of data subjects with relevant protected characteristics under the PSED. This could help address the concern that privacy and data protection laws are too individualistic in their structure to address downstream, societal harms of data processing.

A 'privacy-based' approach to justification therefore helps to expand and fortify data protection as a tool to regulate secondary uses of patients' data. Where patients' identifiable data are used for secondary purposes, these uses should be presumed to engage Article 8, and require justification under Articles 8(2) and 14 ECHR. This means that the GDPR's obligations, such as the DPIA requirement,[193] the principle of 'data protection by design'[194] and the scientific research exemptions[195] should be interpreted in a way which gives due weight to the ECHR requirements of proportionality and non-discrimination. The overarching contribution of Part 3 of my thesis is thus to strengthen the GDPR's safeguards within secondary uses of patients' data.

[193] Ibid.
[194] GDPR, Article 25.
[195] Set out in Article 89, GDPR. These modify most of the GDPR's principles in Article 5 (e.g. transparency, purpose limitation and storage limitation), and the majority of the data subject rights (Articles 13-21). This is discussed further in the 5.academic governance article, **thesis page 75**.

Publication 5

Publication Title: Governance Of Academic Research Data Under the GDPR— Lessons from the UK

Short Title: ‘Academic Governance Article’

Page Number of Thesis: 75

The following pages incorporate the published PDF of this journal article, as it has been made available online on an open-access basis. The formatting, pagination and footnote numbering of the publication have been retained, and thus the numbering is self-contained.

The pagination sequence of this thesis will then resume at the end of the publication.

Governance of academic research data under the GDPR—lessons from the UK

Miranda Mourby*, Heather Gowans*, Stergios Aidinlis*, Hannah Smith* and Jane Kaye*

Key Points

- The General Data Protection Regulation (GDPR) includes a new power for Member States to pass exemptions for the purpose of 'academic expression'.
- This may appear to provide greater freedom to researchers working under the new EU data protection regime.
- Using the UK as a case study, however, it is evident that even a full exercise of the academic derogation is likely to be limited by the GDPR's requirement of necessity, and by privacy rights wherever they are engaged.
- Ultimately, the GDPR provisions applicable to universities as public authorities are likely to have greater impact on academic data processing in public institutions; a shift which is not conducive to greater freedom in research data processing.

Introduction

The General Data Protection Regulation (GDPR or 'the Regulation')[1] provides EU Member States with a new power to make exemptions for academic expression. The obvious question which arises, and which this article aims to address, is whether academics processing data for research therefore stand to have greater freedom under the GDPR?

We answer this question by using the UK as a case study. The UK has made generous use of what we will call the 'academic derogation' in Article 85 GDPR, and as such provides a useful example of what academic exemptions could look like at their fullest extent. While some cross reference is made to jurisdictions such as Malta, Ireland, and Austria which have passed similar national laws, the main focus of this article is on the GDPR and the UK exercise of its research-related derogations. This national example illustrates the intersection of data protection and human rights law which both gives rise to, and limits, academic freedom in data processing.

The second part of this article examines the academic derogation itself. We outline how the UK has exercised the Article 85 derogation to create academic exemptions, and consider how the GDPR's requirement that these exemptions be made only where '*necessary*' to reconcile freedom of expression and information with the right to data protection is likely to limit their scope in practice. Third and fourth part examine other GDPR derogations which may impact upon academic research data: namely the research derogation in Article 89, and the potential for future derogation under Article 23.

The common thread running through our review of these GDPR derogations is the way in which the scope of any resulting exemptions is analysed. Scope, we suggest, encompasses not only the potential *breadth* of application (ie the number of scenarios in which an exemption could hypothetically be claimed) but also the *threshold* (the test an academic researcher would have to satisfy to successfully claim the exemption in practice).

* Centre for Health, Law and Emerging Technologies ('HeLEX') University of Oxford
The authors would like to thank Associate Professor Mark Taylor for his comments on this article.
The work leading to this publication has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.
The authors were also supported by the Economic and Social Research Council grant number ES/L007452/1 for the Administrative Data Service. The funders played no role in the writing of this article, or the decision to submit it for publication.

1 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), OJ 2016 L 119/1 (hereafter cited as GDPR).

Taking the 'vertical' and 'horizontal' axes of scope into account, it is evident that the threshold of the exemptions is set high either by restrictions in the GDPR itself, or by the privacy rights which apply across the European Union.

Finally, the last section considers the delineation the GDPR makes between private and public sector organizations. As the UK explicitly designates universities as public authorities for the purposes of the GDPR, it provides a clear example of the way in which the 'bundle of obligations'[2] reserved for public authorities under the GDPR may impact upon academic processing. We focus in particular on the role of consent as a basis for processing in research, and the concern expressed in the GDPR as to whether such consent can be 'freely given' to a public authority. It is this concern for the freedom and autonomy of data subjects, as part of the public authority obligations in general, which we suggest constitutes the norm for academic research data processing under the GDPR. As such, it seems the Regulation will for the most part bring about more new scrutiny than new freedom for academic researchers.

The academic derogation

There are a number of reasons why the derogation in Article 85 GDPR may appear to create greater freedom for researchers. Most obviously, there is its provision of a new power for Member States to pass exemptions from the GDPR for the purposes of academic expression. The terms on which this power is provided, including reference to the right to freedom of expression '*and information*' which the Regulation instructs Member States to reconcile with data protection, are an additional factor. Finally, the term 'academic expression' is inherently (and perhaps inevitably) open-ended, creating a broad range of potential application.

This section analyses these apparent sources of greater academic freedom, before considering the extent to which the threshold requirement of 'necessity' limits the likely scope of the freedom in practice. Beginning with the GDPR itself, Article 85 gives Member States the power to derogate from the Regulation, wherever such deviation is necessary to:

> reconcile the right to the protection of personal data pursuant to this Regulation with the right to freedom of expression and information, including processing for journalistic purposes and the purposes of academic, artistic or literary expression.

The potential for derogation from the GDPR for these purposes is significant, encompassing all Articles within nine out of the 11 chapters of the Regulation.[3] These chapters include Chapters II (principles), III (data subject rights), IV (obligations of data controller and processors) and V (transfer of data to third countries or international organizations), thus covering the key obligations to which a data controller would otherwise be subject. This is, however, broadly consistent with the previous provision for special purposes derogation within Directive 95/46 EC ('the Directive').[4]

For our present purposes, the most important innovation in Article 85 GDPR (as compared to its predecessor in the Directive) is the addition of 'academic' to the special purposes. Otherwise, the special purpose derogation is similar to the Directive in that it does not discriminate between journalistic, academic, artistic or literary purposes, however different these operations may be in practice. Member states are given free rein to derogate for any or all of these purposes, equally or not as the case may be.

In one sense, Article 85 GDPR appears more liberal than the equivalent provision it replaces in the Directive. The Directive required that the processing be '*solely*'[5] for the special purposes, and the derogations to be:

> 'necessary to reconcile the right to privacy with the rules governing freedom of expression.'[6]

The GDPR, on the other hand, requires reconciliation of:

> 'the right to the protection of personal data pursuant to this Regulation with the right to freedom of expression and information.'

The change may be subtle, but it suggests Member States are no longer obliged to take into account privacy rights as a whole, but rather the right to data protection as embodied in the GDPR itself, when passing national exemptions from the Regulation. As it has been argued

2 Oliver Butler, 'Obligations Imposed on Private Parties by the GDPR and UK Data Protection Law: Blurring the Public-Private Divide' (2018) 24 European Public Law 555, 572.

3 GDPR, art 85(2).

4 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, OJ 1995 L 281/31 (hereafter cited as 'the Directive').

5 Although, as David Erdos points out, the word 'solely' has been retained in Recital 153. The suggestion of exclusivity of purpose is therefore not entirely absent from the Regulation. 'Beyond "Having a Domestic"? Regulatory Interpretation of European Data Protection Law and Individual Publication' (2017) 33 Computer Law and Security Review 275, 290.

6 Directive, Art 9.

that the right to data protection differs in nature and scope from the right to privacy,[7] this transition may appear significant. As we demonstrate in the next section, however, this does not mean the right to privacy no longer impacts upon data processing in an academic context; simply that considerations of privacy are not as prominent within by the GDPR itself.

Additionally, the above-cited text of Article 85 GDPR requires Member States to consider the right to freedom of expression '*and information*', potentially adding a further dimension, and greater weight, to the principles in favour of derogation. The new drafting is a fuller reflection of this right as articulated in Article 10 of the European Convention on Human Rights (ECHR),[8] the principles of which are effectively transposed into EU law via the Charter of Fundamental Rights ('the Charter')[9] in the intervening years between the Directive and the Regulation. It would be a natural consequence of this shift for a slightly different balance to be struck under the GDPR, even in spite of the importance the Regulation otherwise affords to the principles of data protection.

Given the new importance attributed to the right to impart information, it is less surprising that the inclusion of 'academic purposes' in Article 85 is another innovation of the Regulation. The addition of academic purposes followed years of discussions, dating back to 2012, with UK funding organizations supporting its inclusion for the benefit of academic research,[10] particularly research which would not be covered by the scientific research exemptions now set out in Article 89 GDPR.[11] The idea of placing journalism and academic social science research on a more equal footing within data protection law was also met with academic support.[12]

Article 11 of the Charter, on which the derogations in Article 85 GDPR are based,[13] reads as follows:

> Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.

It seems novel that a public authority such as a university should require legal protection from interference from other public authorities in its free expression. It appears almost to posit two diametrically contrasting views of academic research: as a service to society carried out under the aegis of a public authority, and as free individual expression requiring protection from the State. This apparent contradiction in the way universities may be viewed under the GDPR may be an illusion, however. Whether these two models of academic research are likely to conflict in practice will depend upon the impact these exemptions will have on academic data governance. We will explore this potential impact with regard to breadth of application, as well as the threshold for application, of the UK academic exemptions.

Scope of the UK academic exemptions

The simplest measure of the breadth of the UK academic derogations is the number of GDPR provisions from which UK data controllers could claim an exemption for such purposes. From this perspective, the DPA 2018's exercise of the derogation offers an impressively comprehensive regime of exemptions. It comprises, *inter alia*:

- a defence to the crime of unlawfully obtaining personal data;[14]
- a defence to the crime of re-identifying de-identified data without the data controller's consent[15] (although this is technically not a derogation from the GDPR but rather from a national provision, recommended by UK's National Data Guardian for Health

7 Orla Lynskey, 'Deconstructing Data Protection: The "added-value" of a Right to Data Protection in the EU Legal Order' (2014) 63 International & Comparative Law Quarterly 569. Juliane Kokott and Christoph Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 222.

8 European Convention for the Protection of Human Rights and Fundamental Freedoms, 3 September 1953, ETS 5, 213 UNTS 221 (hereafter cited as ECHR).

9 Charter of Fundamental Rights of the European Union, OJ 2010 C 83/389 (hereafter cited as 'EU Charter of Fundamental Rights').

10 Economic and Social Research Council, 'Response to the European Commission's proposed European Data Protection Regulation (COM (2012) 11 final)' (21 February 2013) <https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-response-to-the-european-commission-s-proposed-european-data-protection-regulation-2013/> accessed 2 October 2018.

11 Wellcome Trust, 'Academic research perspective on the European Commission, Parliament and Council texts of the proposal for a General Data Protection Regulation – 2012/0011 (COD)' (2015) <https://wellcome.ac.uk/sites/default/files/research-perspective-data-protecton-regulation-proposal-wellcome-jul15.pdf> accessed 2 October 2018.

12 David Erdos, 'Freedom of Expression Turned On Its Head? Academic Social Research and Journalism in the European Privacy Framework' (2013) 1 Public Law 52.

13 The draft of the GDPR proposed by the European Parliament in March 2014 explicitly refers to the Charter in what was then art 80, but this reference was removed in the text put forward by the Council of Ministers in June 2015. The Charter is not mentioned in the final version of Article 85 of the GDPR, but its text, and that of Article 10 of the European Convention on Human Rights from which Article 11 derives, is still reflected in the wording 'right to freedom of expression and information'. The 'right to freedom of expression and information, as enshrined in Article 11 of the Charter' is still referenced in Recital 153 GDPR, which elaborates on the special purposes.

14 Data Protection Act 2018 (UK) (hereafter cited as DPA), s 170(3)(c).

15 Ibid, s 171(4)(c).

and Care as a protection for anonymized patient data used in health research);[16]

- a defence to the crime of processing data which has been illegally re-identified;[17]
- an additional ground for processing special categories of personal data in order to uncover dishonesty or mismanagement;[18]
- exemptions from all GDPR data subject rights (except the right not to be subject to significant decisions based solely on automated processing), and all the principles relating to the processing of data except for accountability and security.[19]

To expand upon the final bullet point, while paragraph 26(9) of Schedule 2 DPA 2018 does not exercise the full extent of the academic derogation, the following table illustrates the extent to which key GDPR principles, rights, and obligations could be curtailed for academic purposes:

Article 85 Derogation	DPA 2018 Academic Exemption
Chapter II (**Principles**)	**All except** Article 5.1(f) (requirement to process data in a manner which ensures appropriate security) and Article 5.2 (the accountability principle). Ie no requirement for the principles of lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, or storage limitation to be applied where the exemption is successfully claimed.
Chapter III (**rights of the data subject**)	**All except** Article 12 (which has little impact without the application of Articles 13–21, is it regulates compliance with these obligations) **and** Article 22 (the right not to be subject to decisions based solely on automated processing).
Chapter IV (**controller and processor**)	**Article 34** (communication of personal data breach to the data subject) & **Article 36** (requirement to consult the Commissioner prior to high risk processing) only.
Chapter V (**transfers of personal data to third countries or international organizations**)	**Article 44** only
Chapter VI (independent supervisory authorities)	None
Chapter VII (cooperation and consistency)	**Articles 60–67** only
Chapter IX (specific data processing situations)	None

It is evidently not fair to say that the UK has exercised the Article 85 derogation to the fullest possible extent, but the exemptions academics *can* claim would counter the majority of principles and data subject rights in the GDPR. All requirements to process data lawfully, including the need to have an appropriate lawful basis for processing, could be overruled. The transparency obligations, one of the cornerstones of the GDPR, could also be negated—along with the need to ensure accuracy and minimization of data.[20]

Significant provisions which are *not* included in the permissible exemptions include those governing the general responsibilities of a data controller,[21] their relationship with a processor,[22] the need to keep records of processing,[23] cooperate with a supervisory authority,[24] and ensure adequate security of processing.[25] Security

16 Explanatory Notes to the Data Protection Act 2018, para 49.
17 Ibid, s 171(7)(c).
18 Ibid, Schedule 1 para 13.
19 Ibid, Schedule 2, para 26 (9).
20 GDPR, Art 5.1(a)-(e), which are included in the list of potential exemptions.
21 Ibid, Art 24.
22 Ibid, Arts 28-29.
23 Ibid, Art 30.
24 Ibid, Arts 31 and 33.
25 Ibid, Art 32.

of processing is also protected by the preservation of Article 5(1)(f). Clearly, the free expression Article 85 GDPR protects is only *deliberate* expression and information through personal data; accidental revelation of personal information through inadequate security would not fall within this protected right.[26] This is an interpretation evidently shared by Malta[27] and Ireland,[28] who have also specifically excluded Article 5(1)(f) from their academic exemptions. Austria has preserved Article 5 in its entirety,[29] but allows more exemptions from data controller obligations.[30]

The greatest concern which arises from a review of the UK exemptions, therefore, relates more to the potential absence of lawfulness, transparency, fairness, data minimization and the availability of data subject rights, than to the structure and organization of the controller's processing. Nonetheless, it will be interesting to see how much the continued effect of Article 35 (the duty to prepare a data protection impact assessment) which is not included within the exemptions, will limit the exercise of the other exemptions. Presumably, even when exemptions *are* claimed from the lawfulness principle, or from data subject rights, the controller will still have to conduct a data protection impact assessment where there is a high risk to individuals' rights and freedoms, and mitigate against these risks wherever possible. Other reasons why necessity and proportionality must be considered are discussed below under 'threshold', but the continued application of Article 35 strongly suggests that exercise of academic exemptions should be planned, and proportionate to the risks posed to data subjects.[31] This is consistent with the guidance provided by the UK Information Commissioner's Office (ICO) that use of all GDPR exemptions should be justified and documented.[32]

'Academic material'

An additional measure of scope is the likely definition of 'academic material': the central concept on which the exemptions hang. Academic purposes are not defined in the GDPR or the DPA 2018, but 'academic material' intended for publication is a prerequisite to disable the GDPR provisions listed in the table above.[33] This, in turn, is never defined within the DPA 2018. The Charter refers to the need to respect 'academic freedom', but does not provide further detail.[34] In the context of the freedom of expression and information, on which the Article 85 derogation is based,[35] some gloss was provided by the Court of Justice of the European Union in the *Erdogan* case:

> In determining whether speech has an 'academic element', it is necessary to establish: a) whether the person making the speech can be considered an academic; b) whether that person's public comments of utterances fall within the sphere of his or her research; and c) whether that person's statements amount to conclusions or opinions based on his or her professional expertise and competence. These conditions being satisfied, an impugned statement must enjoy the utmost protection under Article 10.[36]

None of these criteria are reflected in the text of the DPA 2018, which adopts an interpretation of the derogations which is 'actor neutral' (ie referring to special purposes and not to special actors).[37] It is possible that the ICO might take them into account when determining whether data are processed 'only' for academic purposes,[38] but the regulator's criteria has yet to be concretized, and as such it is difficult to speculate at this stage. This is not necessarily a deficiency in the DPA 2018, however. It would be difficult to be too prescriptive about the breadth of academic material, and what constitutes valid information, opinions or ideas. It would be almost paradoxical for parliament to attempt to safeguard freedom of expression, while at the same time attempting to impose its own definition of legitimate academic discourse.

While the breadth of 'academic material' is not necessarily a deficiency in the DPA 2018, it nonetheless creates scope for confusion or controversy between academics and their institution, particularly as regards the identity of the data controller of the exempted data. Even taking the *Erdogan* gloss into account, all that would be required for 'academic material' is an individual who could fairly be termed an academic, a connection between said individual's field of research and the information in question, and a conclusion or opinion

26 This is consistent with the previous exemptions for journalists and artists under the UK's Data Protection Act 1998, s.32.

27 Data Protection Act, Cap 586 28 May 2018 (Malta), s 9(2)(a)(i).

28 Data Protection Act 2018, Number 7 of 2018 (Ireland), s 43(2).

29 Data Protection Amendment Act 2018 (*Datenschutz-Anpassungsgesetz 2018*) 31 July 2017 (Austria), s 9.

30 All of Chapter IV except from arts 28, 29 and 32 is covered by the academic exemptions (see note 29 above).

31 GDPR, art 35.7(b)–(d).

32 Information Commissioner's Office, 'Guide to the General Data Protection Regulation/ Exemptions' <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/exemptions/> accessed 9 October 2018.

33 DPA, Schedule 2, para 26(2).

34 EU Charter of Fundamental Rights, art 13.

35 (n 13).

36 *Mustafa Erdogan and Others v Turkey* App nos 346/04 and 39779/04 (ECtHR, 27 May 2014).

37 Erdos (n 5).

38 DPA, s 174.

based on their professional competence. This is too broad a principle for it to be clear whether the 'academic material' needs to be held or controlled by an academic institution. For example, in the well-publicised case of Aleksandr Kogan's use of Facebook data (the results of which allegedly came into the possession of Cambridge Analytica) the University of Cambridge was clear that no University resources or facilities, and none of the data collected for his academic research, were used for this work.[39] This was sufficient for the UK regulator to focus their investigation on Dr Kogan's private company, rather than his academic employer.[40] If a researcher in a similar position were to process data under DPA 2018, outside of their academic employment but using their professional competence to draw conclusions intended for publication, there is nothing in the foregoing analysis to suggest they would not be entitled to rely on academic exemptions, even relating to fundamental data protection principles such as that of lawfulness. The DPA 2018 thus contains a significant potential lacuna in respect of 'spin out' academic processing, although (as argued in the next subsection) the continuing application of privacy rights may be a saving grace for affected data subjects.

There is also nothing within the DPA 2018 to prevent an academic who processes data within the course of their employment from moving institution and taking their exempted data with them, as the purpose of the academic material could move with the researcher. The phrasing in the *Erdogan* judgment focuses on the academic expertise of an individual, and even the DPA 2018's '*with a view to publication*' could refer to the intention of a particular individual to publish the results of their data processing. That said, nothing within data protection law is capable of bestowing intellectual property rights on an academic. Even if said individual were to leave a university, taking their professional competence and intention to publish with them, the DPA 2018 could not assist if the terms of their employment meant intellectual property rights in their research data remained with their university. An interesting situation might arise if the university could not replace this individual with someone with similar competence and intention to publish—in this instance the university might retain IP in the data, but lose the basis on which to rely on any academic exemptions under DPA 2018.

If the potential scope of the exemptions is broad, albeit perhaps justifiably, it is all the more important that an appropriate threshold is set to prevent abuse of academic freedom. This is considered in the next subsection.

Threshold for the academic exemptions

Starting with the test as set out in the DPA 2018 itself, the criteria a controller would have to satisfy under can be summarized as follows:

- Regardless of the flexibility in Article 85 GDPR, under UK law the personal data must be processed *only* for academic purposes, this being a point the ICO has power to determine;[41]
- Academic purposes constitute any processing with a view to the publication by a person of academic material which has not been previously published by the controller;[42]
- The controller must reasonably believe that the publication of this material would be in the public interest;[43] taking into account the importance of the public interest in the freedom of expression and information[44] and any one of the following:
  - BBC Editorial Guidelines
  - Ofcom Broadcasting Code
  - Editors' Code of Practice

to the extent that these Codes are relevant to the publication in question;[45]

- The GDPR provision from which the controller wishes to claim an exemption be (in their reasonable belief) incompatible with the academic purposes.[46]

These criteria are spread across the DPA 2018, and can pose a logical puzzle in their assembly. For example, paragraph 26(2) of Schedule 2 DPA 2018, sets the test for when the listed GDPR provisions can be said not to apply:

> (2) Sub-paragraph (3) applies to the processing of personal data carried out for the special purposes if—

39 University of Cambridge, 'Statement from the University of Cambridge about Dr Aleksandr Kogan' 23 March 2018 <https://www.cam.ac.uk/notices/news/statement-from-the-university-of-cambridge-about-dr-aleksandr-kogan> accessed 6 May 2019.

40 Information Commissioner's Office, 'Investigation into the use of data analytics in political campaigns' report to Parliament 6 November 2018 <https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf> accessed 6 May 2019.

41 DPA, s 174(3)(a).

42 Ibid, ss 174(3) and 176(1).

43 Ibid, Schedule 2, Part 5, para 26 (2) (b).

44 Ibid, Schedule 2, Part 5, para 26 (4).

45 Ibid, paras 26 (5) and (6).

46 Ibid, para 26 (3).

(a) the processing is being carried out with a view to the publication by a person of journalistic, academic, artistic or literary material, and
(b) the controller reasonably believes that the publication of the material would be in the public interest.

This does not include any requirement that academic material be previously unpublished, or that processing should be '*only*' with a view to such publication. At the same time, however, if a data subject were to make a claim under Article 82 GDPR (for compensation as a result of infringement of the Regulation), the ICO could make a determination within those proceedings that the personal data are not processed '*only*' for the special purposes, or for the sake of previously unpublished material.[47] If the processing *were* only for the sake of previously unpublished academic material, such proceedings must be stayed,[48] but otherwise the controller could presumably be held liable for the processing. Furthermore, it seems from the face of the Act that 'with a view to' means that the processing must *anticipate* publication, meaning the exemption would no longer be available post-publication. This is less appropriate in an academic context as it might be within journalism, as academics may be required by journals to retain data post publication for validation purposes. They may even be obliged to share said data by public funders. Although the 'archiving in the public interest' exemptions might offer some assistance in this context, we shall demonstrate in Part III that these 'Article 89' exemptions are far more narrow than those afforded for academic processing. The academic exemption relied upon might therefore disappear post-publication. Therefore, it would be foolhardy for a university to rely on the text of Schedule 2 paragraph 26 alone, and the totality of the DPA 2018 must be taken into account.

The threshold for the academic exemptions set by the DPA 2018 is thus complex in the sense that it must be read cumulatively across the Act. It is not necessarily more liberal than its counterparts: the Irish Data Protection Act 2018 does not appear to have a 'sole purpose' requirement, and explicitly states that the right to freedom of expression and information must be interpreted in a broad manner.[49] The Maltese Data Protection Act 2018 is interesting, however, in adding:

> Personal data processed for the purpose of exercising the right to freedom of expression and information, including processing for journalistic purposes or for the purposes of academic, artistic or literary expression, shall be exempt from compliance with the provisions of the Regulation specified in sub-article (2) where, having regard to the importance of the right of freedom of expression and information in a democratic society, compliance with any of the provisions as specified in sub-article (2)would be incompatible with such processing purposes:
>
> **Provided that** when reconciling the right to the protection of personal data with the right to freedom of expression and information, **the controller shall ensure that the processing is proportionate, necessary and justified for reasons of substantial public interest**[50] (emphasis added).

This language appears to reflect the requirements for lawful interference with a fundamental right, as though the legislation anticipates the possibility that data processing under the special purposes exemptions would engage data subjects' privacy rights. While it cannot be assumed that processing personal data for academic research will always engage such rights, information of a private nature will be processed with sufficient regularity in the course of research (particularly social or biomedical scientific research) that this is an important factor to consider, which could significantly restrict the scope of the exemptions in practice.

Privacy rights

Although considerations of necessity and proportionality are not explicitly written into the test for the academic exemptions in the DPA 2018, there are reasons outside the Act itself why these principles should be read into their application in practice.

Firstly, even though the exemptions are the UK's reconciliation of the *right to data protection* with freedom of expression, the right to *privacy*[51] as expressed in Article 8 ECHR applies directly in the UK by virtue of the Human Rights Act 1998, which creates a cause of action distinct from data protection law. That is, even if no infringement of data protection regulation had taken place, a data subject could still theoretically bring a claim for breach of their rights under Article 8 ECHR (even if the two are in fact found to align closely in practice).

The Human Rights Act 1998 alone provides grounds to assert that university researchers should respect privacy rights.[52] These rights apply directly to universities

47 Ibid, s 174(3).
48 Ibid, s.176.
49 Ibid, s 43.
50 Ibid, s 9.
51 That is, the right to 'respect for private and family life, home and correspondence'.'
52 Deryck Beyleveld and others, 'The UK's Implementation of Directive 95/46/EC' in Deryck Beyleveld and others (eds), *Implementation of the Data Protection Directive in Relation to Medical Research in Europe* (Routledge, London 2018).

if they are public authorities for administrative law purposes.[53] Even if a university is taken to be a private organization, Article 8 ECHR applies indirectly, as it will be brought to bear on the interpretation of their private law duty of confidentiality.[54]

Secondly, Article 8 ECHR will influence how the GDPR and supplementing national legislation are interpreted, including the academic exemptions under the DPA 2018. As long as the GDPR applies—and UK's legislation merely supplements its provisions—this framework should be interpreted in line with fundamental rights set out under the Charter, which largely mirror those of the ECHR. This was settled case law under the Directive,[55] and it would be difficult to argue that it should be otherwise under the GDPR given the prominence the Regulation gives to the Charter. The right to data protection as enshrined in the Charter is referred to in the first Recital of the GDPR, setting the tone of the rest of the Regulation. Recital 4 GDPR also adds:

> This Regulation respects all fundamental rights and observes the freedoms and principles recognised in the Charter as enshrined in the Treaties, in particular the respect for private and family life, home and communications.

It therefore seems in no way far-fetched to argue that the GDPR should be interpreted in line with the Charter, as a continuation of the principle established under the Directive. The UK Supreme Court has gone as far as to suggest that separate consideration of the right to data protection was not necessary as:

> In Volker und Marcus Schecke GbR and Hartmut Eifert v Land Hessen (Cases C-92/09 and C-93/09) [2010] ECR I-11063, the Court of Justice of the European Union (Grand Chamber) held (para 52) that the limitations which may lawfully be placed on the right to the protection of personal data correspond to those tolerated in relation to article 8 of the ECHR.[56]

This equation has been followed in the lower courts, where it has been recently accepted that a claim under the Data Protection Act 1998 would add nothing to one under Article 8 ECHR.[57] As the right to privacy in Article 7 of the Charter is expressed in terms substantively identical to Article 8 ECHR, it is convenient to refer to both provisions collectively as 'privacy rights'.

This is not to accept a characterization of all GDPR subject rights as equating to Article 8 ECHR rights. Despite the tendency of courts to often equate data protection and privacy rights in their analysis,[58] there are reasons for avoiding a conflation between the two.[59] The rights to access or rectification under Articles 15 and 16 GDPR, for instance, correspond to a notion of privacy-as-control that applies regardless of a potential 'interference with private life'.[60] Nonetheless, we shall be cautious before drawing too rigid a distinction between these different elements of data protection law. All data subject rights could be exempted from for academic purposes, unless such processing was to interfere disproportionately with private life, and thus the precise scope of an academic exemption *vis a vis* a data subject right, is largely contingent on the specific factual context.[61]

For all of these reasons, it is clear that the academic exemptions must be construed in line with privacy rights wherever private information is used in research. The inevitable question is then how these rights limit the scope of the exemptions? The GDPR limits the academic derogation to that which is 'necessary' to reconcile the right to freedom of expression and information with data protection. It therefore follows that the exemptions should go no further than that which is 'necessary' for this reconciliation. As the UK Supreme Court explained in the above-cited case:

> The meaning of necessary was considered by this court in *South Lanarkshire Council v Scottish Information Comr* [2013] UKSC 55; 2014 SC (UKSC) 1; [2013] 1 WLR 2421. As was explained there at paras 25-27, it is an expression whose meaning depends on the context in which it falls to be applied. Where the disclosure of information constitutes an interference with rights protected by article 8 of the ECHR, as in the present context (as explained at paras 75-77 below), the requirement that disclosure is "necessary"

53 There is some authority to suggest universities may be 'public authorities' from an administrative law perspective in their educational capacities, but not in their private capacity as employers: *Evans v University of Cambridge* [2002] EWHC 1382 (admin); *R (Galligan) v University of Oxford* [2001] EWHC Admin 965. As such, it is possible the Human Rights Act 1998 would apply to universities as public authorities in their research functions.

54 *Campbell v Mirror Group Newspapers Limited* [2004] UKHL 22, at para 17.

55 Case C274/99 P *Connolly v Commission* [2001] ECR I-1611, at para 37; and Case C-465/00 *Rechnungshof v Österreichischer Rundfunk* [2003] ECR I-4989, at para 68.

56 *Christian Institute v Lord Advocate (Scotland)* [2016] UKSC 51, at para 104.

57 *Richard v BBC* [2018] EWHC 1837 (Ch), at para 226.

58 Cases C-92 and 93/09 *Volker und Marcus Schecke GbR and Hartmut Eifert v Land Hessen* EU:C: 2010:662 at para 52; *NT1 v Google LLC* [2019] QB 344, at para 9 (Warby J).

59 (nn 8 and 9).

60 Gloria Gonzalez Fuster, *The Emergence of Personal Data Protection as a Fundamental Right of the EU* (Springer 2014) 205.

61 Also see the ECtHR's context-specific assessment of the potential conflict between the right to private life and the freedom of journalistic expression in *Magyar Helsinki Bizottsag v Hungary, App no 18030/11* (Judgment of 8 November 2016).

> forms part of a proportionality test: the disclosure must involve the least interference with the right to respect for private and family life which is required for the achievement of the legitimate aim pursued.[62]

Even where no 'disclosure' is involved, and data are merely processed for research, we suggest the principle is the same. If privacy rights are engaged by research data processing, this processing should not exceed that which is necessary for the ultimate academic publication. This appears to correlate best with the requirement in the DPA 2018 that there be incompatibility between the exempted data protection obligation and the academic purpose. In other words, wherever privacy rights are engaged, 'incompatibility' should be read as meaning: 'is it strictly necessary to deviate from the GDPR, or could the academic goal be achieved with a less intrusive (and more compliant) measure?'

Therefore, although there is no explicit requirement within the DPA 2018 to weigh the perceived public interest in the ultimate publication of the material against the severity of the effects on the data subject's right to privacy,[63] these considerations must instead be prompted by reading the GDPR in line with privacy rights.

A hypothetical scenario outlining when Article 8 ECHR might impose additional considerations beyond these criteria is set out below, in an example which also illustrates the potentially complex relationship between the different types of exemption:

> A group of scientific researchers is processing a large amount of sensitive research personal data for their project, the findings of which are believed to be in the public interest and are intended for publication. They did not obtain the data directly from the data subjects, but from another centre as part of a data sharing consortium. These researchers have not actively communicated all of the information listed in Article 14 GDPR directly to the individual data subjects, believing that to do so would require a disproportionate effort, and would render their research impossible as it would exhaust their budget.
>
> They note that they could rely on Article 14. 5(b) of the GDPR, and not inform data subjects on the basis of the disproportionate effort involved, but that to do so they would need to have in place safeguards pursuant to Article 89 GDPR. They are concerned about the time and resources these safeguards might involve, and wonder whether this burden might mean the safeguards are 'incompatible' with their purposes (as required by the test in the DPA 2018)? Could they not simply rely on DPA 2018 Schedule 2, para 26(9)(b)(ii), which would exempt them from compliance with Article 14 altogether, and not put in place any safeguards? They could also use similar DPA 2018 exemptions to avoid compliance with other provisions which require Article 89 safeguards, such as the 'research' condition for processing special categories of data under Article 9.
>
> Whether the safeguards Article 89 requires could truly be said to be 'incompatible' with the academic purposes, as opposed to purely 'inconvenient',[64] will depend on the circumstances of the case. However, case law from the European Court of Rights suggests that where large amounts of sensitive data are processed, safeguards protecting these data from abuse are all the more important for compliance with Article 8 ECHR.[65] As such, to avoid infringing privacy rights, a more robust interpretation of 'incompatible', as used in the DPA 2018, may be required.
>
> As such, the researchers may be able to rely on the scientific research exemption, meaning they could rely on Article 14. 5(b) GDPR and only make information about their processing publicly available, rather than incurring the expense of contacting data subjects individually. However, to try to dispense with research data safeguards on the basis that they would be 'incompatible' with their purposes sets the bar for incompatibility too low to be compatible with privacy rights. Therefore, it is unlikely that the academic exemptions in DPA 2018 could be used to avoid putting in place Article 89 safeguards to protect private information.

Following the above analysis, it is suggested that the principle of proportionality should be read into the word 'incompatible' in practice, wherever private information is used. Closer analysis reveals a key distinction between the academic exemption and the additional consideration of privacy rights. Under the academic exemption, the researcher must only have a 'reasonable belief' of incompatibility. As a potential arbiter of any infraction, the ICO confirms:

62 *Christian Institute* (n 56) at para 56.

63 The ICO's guidance (n 40) suggests data controllers should consider the harm to data subjects when deciding whether the ultimate publication would be in the public interest. We endorse this guidance, but note that it is not a requirement of the DPA 2018 itself.

64 The Information Commissioner's Office has since released guidance confirming that incompatibility must be more than inconvenience: 'Guide to the General Data Protection Regulation/ Exemptions' (note 40)

65 *MM v UK* App no 24029/07 (ECtHR, 13 November 2012), at para 200.

> The ICO does not have to agree with your view – but we must be satisfied that you had a reasonable belief.[66]

Likewise, the courts would be likely to apply a 'reasonable range' test and afford some discretion for reasonable belief when considering an exercise of the special purposes exemptions. In the case of an alleged infringement of privacy, however, the court would make its own determination of proportionality, with no allowance for reasonable alternative views. As such, a greater degree of external accountability is established once privacy rights are taken into account, as the extent to which these rights can be impaired is ultimately for the courts to determine.

As a counterexample, the following scenario outlines a situation in which the academic exemptions could supplement the research exemptions, and broaden the flexibility for university researchers:

> A study has obtained data directly from research participants, spending a significant amount of public money on this initial individual level engagement. The information originally provided to research participants was deemed sufficient under the applicable data protection law at the time of collection. However, now that the GDPR is in force, Recital 171 requires the study to bring their research personal data in line with the Regulation. This means imparting additional information to data subjects in order to satisfy the GDPR's enhanced transparency requirements, such as the legal basis on which their data are processed, and the availability of their GDPR rights. Where such information goes to the heart of how data subjects exercise their rights, European level advice is that it should be actively brought to the attention of each data subject, and not simply advertised on a website.[67] The study does not, however, have the resources to go back to the thousands of data subjects to make good this informational deficit, especially as the budget for the study has been all but exhausted.
>
> The analysis of the personal data for the originally stated purpose has not yet completed, and nothing about the way the data are processed has changed. The researchers are satisfied there is no significant impact on the participants, as their personal data are used only in line with their original consent. But compliance with Article 13 to the extent of direct contact with each participant would exhaust the funding, and compromise the ability of the researchers to generate the findings for which the data were originally obtained. This would jeopardise the research for which participants had given up their time and information in the first place.
>
> As the data are intended to generate findings for publication believed to be in the public interest, but a requirement to bring the data fully in line with Article 13 would mean the end of the project and the deletion of the data, in this extreme case the UK academic exemptions may be of some utility.

In summary, therefore, while the academic exemptions could apply in a number of scenarios, genuine incompatibility of the GDPR with academic processing is not easy to make out, especially when considered in the context of the duty to respect privacy rights wherever they are engaged by research. The threshold of necessity within Article 85 GDPR is high, particularly when read in line with privacy rights. As such, the impact of the exemptions is likely to be limited in practice.

However, the academic derogation is not the only means by which the GDPR allows member states to pass exemptions for academic research. For scientific, historical and statistical research, the national exemptions enacted under Article 89 GDPR are of equal significance. While these exemptions are of more limited disciplinary scope, and are less radical in their interference with GDPR obligations, they may be more commonly (and less controversially) deployed within academic research. This will be explored in the next section.

The research derogation

The derogation in Article 89(2) GDPR is another respect in which national implementation of the Regulation can impact upon academic research data. Even where academic exemptions are not claimed, universities can still rely on modifications within the research exemptions.

There are a number of contexts in which the GDPR permits such adaption of its general obligations: for scientific or historical research, for statistical purposes, or for archiving in the public interest. Article 89(2) provides:

> '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

66 Ibid, n 40.

67 Art 29 Working Party, 'Guidelines on transparency under Regulation 2016/679' (WP 260, 11 April 2018), at 5.

render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.

Article 89(1), in turn, refers to '*appropriate safeguards*' for the rights and freedom of the data subject, further specifying:

'Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner.'

Article 89, and its requirement of safeguards for the rights and freedoms of data subjects, is thus the gateway to these research modifications of GDPR obligations. Compliance with Article 89 is one condition on which the general prohibition on processing special category data can be set aside for scientific or historical research;[68] and can enable less active compliance with the transparency obligations where data are obtained from a third party.[69] It can also lengthen permitted retention periods,[70] and in some circumstances can exempt data controllers from the application of data subject rights (see below).

As successful compliance with Article 89, and its clear (if broad) requirement of 'appropriate safeguards', is necessary for these exemptions, we shall refer to them collectively as 'the research exemptions.' The Article 89 derogation is less fundamental in its scope than the academic derogation, as it does not affect the essential requirements of fair and lawful processing. It is not, for example, a substitute for a legal basis for processing. Nonetheless, it can alter a data controller's obligations in ways which are not insignificant. Article 89 GDPR has been criticized for its broad definition of 'scientific research', and for the vagueness of its key term: 'appropriate safeguards'.[71] It is evidently a derogation which, despite the potential for more regulation of research data publication in earlier drafts,[72] does not provide a detailed framework for research data protection, but rather flexibility for research and archiving, as long as sufficient safeguards are in place.

For our present purposes, the most pertinent aspect of Article 89 is the potential for member states to derogate from certain data subject rights for relevant research,[73] or for archiving in the public interest.[74] This has implications both for research itself and for research data archiving at national level. The UK has exercised these derogations in respect of all possible data subject rights in Schedule 2 DPA 2018. Once the exemptions for scientific and historical purposes from both the GDPR and the DPA 2018 have been applied, these rights are modified as follows:

Data Subject Right	GDPR Modification	DPA 2018 Exemption
Access to information about personal data processing (Article 15)	None	Derogation exercised—does not apply to the extent the right would prevent or seriously impair the achievement of the processing purposes.[75]
Rectification (Article 16)	None	Derogation exercised—does not apply to the extent the right would prevent or seriously impair the achievement of the processing purposes.[76]
Erasure (Article 17)	Does not apply to the extent the right is likely to render impossible or seriously impair the processing objectives.[77]	No derogation available—exemption already exists within the GDPR.
Restriction (Article 18)	If exercised, data can only be processed with the data controller's consent, except for reasons of 'important public interest'—not specific to the research exemptions.[78]	Derogation exercised–Does not apply to the extent the right would prevent or seriously impair the achievement of the processing purposes.[79]
Notification obligation regarding rectification, erasure or restriction (Article 19)	Applies unless compliance would be impossible or involve disproportionate effort[80]	Derogation exercised by DPA 2018 — power to derogate only exists for archiving under Article 89(3), and not for research under Article 89(2).[81]

Continued

68 GDPR, art 9(j).
69 Ibid, art 14(5)(b).
70 Ibid, art 5(1)(e).
71 Kart Pormeister, 'Genetic Data and the Research Exemption: Is the GDPR Going too Far?' (2017) 7 International Data Privacy Law 137.
72 Nikolaus Forgo, 'My Health Data—Your Research: Some Preliminary Thoughts on Different Values in the General Data Protection Regulation' (2014) 5 International Data Privacy Law 54.
73 GDPR, art 89(2).
74 GDPR, art 89(3).
75 DPA, Schedule 2, Part 6, paras 27 (2)(a) and 28(2)(a).
76 Ibid, paras 27(2)(b) and 28(2)(b).
77 GDPR, art 17.3(d).
78 Ibid, art 18.2.
79 DPA, paras 27(2)(c) and 28(2)(c).
80 GDPR, art 19.
81 DPA, para 28 (2)(d).

Continued

Data Subject Right	GDPR Modification	DPA 2018 Exemption
Data Portability (Article 20)	None, although the right does not apply to the extent the processing is necessary for a task in the public interest, and only arises if the basis for processing is consent.[82]	Derogation exercised by DPA 2018 — power to derogate only exists for archiving under Article 89(3), and not for research under Article 89(2).[83]
Objection (Article 21)	The data subject has the right to object, unless the research processing is necessary for a task carried out for reasons of public interest.[84]	Derogation exercised by DPA 2018—does not apply to the extent the right would prevent or seriously impair the achievement of the research purposes.[85]
Right not to be subject to significant decisions based solely on automated decision making (Article 22)	Applies except where decision is authorized by Union or Member State law[86]—unlikely to apply within research.	No derogation available.

The above table may appear complex—such is the intricacy of the interface between the two pieces of legislation—but the net result of setting them side by side is to illustrate that the UK has exercised the Article 89 derogation to the fullest possible extent. This suggests a desire to provide researchers, and research archivists, with the greatest possible degree of freedom which is, nonetheless, consistent with the GDPR. The UK is not unique in this regard: Ireland[87] and Malta[88] have also exercised the Article 89 derogation to the fullest possible extent.

At first glance, it may appear troubling that almost all GDPR data subject rights are derogated from under the UK's implementation of the research exemptions. Furthermore, the main criterion for compliance with Article 89—'appropriate safeguards—is non-prescriptive and open to national interpretation. The only specification the UK provides as to what constitutes an 'appropriate safeguard' for scientific research relates to processing for decisions with respect to the data subject, or which are likely to cause substantial damage or distress.[89] Otherwise, 'appropriate safeguards' are left to largely individual data controller discretion, albeit with some particular emphasis on data minimization. The UK has evidently taken a liberal approach in its implementation of the research exemptions; an approach which could impact upon almost all of the rights which would otherwise be available to research data subjects.

However, there is a note of reassurance which has some parallel with the academic exemptions. If the scope is correctly understood as encompassing both breadth and threshold, it can be seen that the threshold built into Article 89 GDPR is high:

> 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**. (emphasis added)

This contrasts with the Directive which only required 'appropriate safeguards' for research derogation, and not serious impairment.[90] The GDPR, however, requires safeguards *and* impossibility or serious impairment for the research derogation, setting a higher standard than existed in the Directive.

In conclusion, therefore, while the GDPR introduces broad derogations which may impact upon academic research, where Articles 85 and 89 are concerned their respective requirements of necessity and (at least) serious impairment limit their application in practice. They are evidently exceptions rather than the norm, unless it is accepted that it is the norm for data protection provisions to seriously impair research (which we do not). As the UK has not allowed any academic exemption from Article 35 (the duty to prepare data protection impact assessments), it is hoped that research data processing will be planned in such a way to promote data subject rights, and accommodate them without serious impairment to research.

On balance, this may appear to be a satisfactory reconciliation of data subject rights, privacy rights, the right to free expression and the importance of public

82 GDPR, art 20(3).

83 Ibid, para 28 (2)(e).

84 Ibid, art 21(6).

85 DPA, Schedule 2, paras 27(2)(d) and 28(2)(f).

86 Or is necessary under a contract, or is based on the data subject's explicit consent: GDPR, art 22(2).

87 Data Protection Act 2018 (Ireland) (n 28), s 61.

88 Data Protection Act, Cap 586 (Malta) (n 27), s 6.

89 DPA, s 19. The prohibition on using research exempted data for decisions or measures relating to an individual has been retained in the GDPR at least as regards statistical purposes (Recital 162).

90 Directive, arts 6(1)(e) & 11(2).

interest research. However, as we note in the next subsection, the number of derogations available under the GDPR means that Member States have significant scope to adapt its provisions, and this may prove a dynamic process.

Future exemptions?

A final GDPR derogation should be considered to complete the picture of academic research governance under DPA 2018: the power in Article 23 for Member States to restrict the scope of certain obligations and rights where necessary to safeguard, *inter alia*, objectives of general public interest.[91] This clearly has implications beyond academic research, and may well have less impact on research governance than Articles 85 and 89 GDPR. Nevertheless, the DPA 2018 'delegates' the power to make regulations under Article 23 to the Secretary of State, thus introducing an element of uncertainty into the Act.

These powers to add to the provisions of the DPA 2018 have been termed 'Henry VIII' powers—otherwise known as delegated legal powers under which subordinate legislation is enabled to amend primary legislation.[92] The inclusion of such powers in primary legislation has been increasingly employed by the UK government, not only to regulate administrative and technical procedures, but also matters of principle and policy.[93] For present purposes, the power in section 16(2) of the DPA is crucial:

> the power in Article 23(1) (GDPR) to make a legislative measure restricting the scope of the obligations and rights mentioned in that Article where necessary and proportionate to safeguard certain objectives of general public interest (may be exercised by way of regulations made by the Secretary of State)

While Henry VIII powers are not entirely a novelty of the DPA 2018, the previous power under the Data Protection Act 1998 was merely to exempt personal data from the subject information provisions where other UK enactments restricted or prevented such disclosure.[94] It was, in essence, a practical power to resolve any conflict between the data subject rights in the Data Protection Act 1998 and any other UK legislation, which still required the Secretary of State to consider the interests of the data subject, as well as rights and freedoms of other individuals.

The power bestowed by section 16(2) DPA 2018 is much broader, enabling the Secretary of State to restrict the scope of any rights or obligations under the GDPR, where such restriction is proportionate and necessary for 'objectives of general public interest'. Unlike its predecessor in the 1998 Act, there is no statutory requirement for the Secretary of State to consider the interests of data subjects. However, although broad, the Secretary's discretion is not untrammelled. Secondary legislation of the kind made under section 16(2) could be subject to judicial review both on traditional *vires* grounds[95] and on the ground of its conformity with fundamental rights. The Supreme Court majority clarified in its recent *Public Law Project* judgement that a 'restrictive' approach is to be preferred if there are doubts about the scope of the delegated power conferred upon the Secretary of State.[96] When fundamental rights are affected by a measure, UK courts employ the so-called 'anxious scrutiny' test, requiring the public authority to prove that 'the most compelling justification existed' for this infringement.[97]

In sum, therefore, although it is not possible to quantify with any certainty the likelihood of the Article 23 derogation being used by the Secretary of State in such a way as impacts upon academic research data governance, this does not appear to be an immediate prospect.

The overarching lesson to be drawn from the UK's response to Articles 85, 89, and 23 GDPR is that the number of available derogations within the Regulation can mean that national implementations are not only complex, with overlapping regimes of restrictions, modifications, and exclusions, but also dynamic and potentially subject to change. Nevertheless, such changes, restrictions and modifications cannot take place in a vacuum. Even in the case of Article 23 where the GDPR does not set thresholds within the derogation itself—

91 GDPR, art 23(1)(e).

92 *R (Public Law Project) v Lord Chancellor* [2016] AC 1531 (Lord Hoffmann), at para 25

93 Hansard Society, 'The Devil is in the Detail: Parliament and Delegated Legislation' (2013) <https://www.hansardsociety.org.uk/publications/the-devil-is-in-the-detail-parliament-and-delegated-legislation> accessed 2 October 2018. The Data Protection Act is no exception to this trend, containing '37 individual regulation-or rule-making powers' as the UK Parliament's Delegated Powers and Regulatory Reform Committee observes, 'Ninth Report of Session 2017-19: HL Paper 48' (6 December 2017). <https://publications.parliament.uk/pa/ld201719/ldselect/lddelreg/48/4803.htm> accessed 2 October 2018.

94 Data Protection Act 1998 (UK), s 38.

95 *Public Law Project* (n 92) (Lord Hoffmann) [26]: 'whether or not it is within the class of action that Parliament must have contemplated when delegating'.

96 Ibid, citing established House of Lords case law in *R v Secretary of State for Social Services Ex p. Britnell* [1991] 1 WLR 198, 204 and *R (on the application of Spath Holme Ltd) v Secretary of State for the Environment, Transport and the Regions* [2001] 2 AC 349, at para 382.

97 *R v Ministry of Defence Ex p Smith* [1996] QB 517, 554; *R (Mahmood) v Secretary of State for the Home Department* [2001] 1 WLR 840.

unlike the requirement of necessity in Article 85, or that of serious impairment in Article 89—they must still be exercised in accordance with case law wherever privacy rights are engaged. The resulting legislative picture is undeniably complex, and not necessarily static, but the impact upon academic research is limited by the need to respect privacy rights, even when this requirement is not explicitly recognized within the GDPR.

Public authority obligations

Finally, we shall now turn to a very different facet of the GDPR: its greater delineation between public and private sector organizations as compared with the Directive it replaces. A review of these two pieces of legislation indicates that the GDPR specifically mentions public authorities 44 times, dwarfing the five such references in the Directive. Included in these references are the requirement for authorities to process data on a national or EU legal basis wherever they exercise their functions, as opposed to relying on the more general latitude of 'legitimate interests'.[98] They must appoint, support and appropriately resource a Data Protection Officer to oversee the lawfulness of their processing,[99] as well as acting as a public-facing point of contact for transparency purposes.[100]

All of these new obligations point towards a greater need for accountability in public authority processing, above and beyond the general enhanced requirements across the public and private sectors. We do not mean to simplify the GDPR's treatment of public authorities, and acknowledge that such institutions also enjoy specific protections, for example from litigation in another Member State or by way of potential national limits on administrative fines.[101] However, as our focus is on academic researchers, and how their relationship with data subjects is regulated under the GDPR, we shall focus on the provisions most pertinent to the question 'to what extent are academics able to process research data as they wish under the GDPR?', as opposed to those pertaining to the consequences of an alleged breach.

To this end, we focus in this final section on consent as a basis for processing data in academic research. The GDPR defines consent as:

> any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.[102]

Recital 43 adds:

> consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority and it is unlikely that consent was freely given in all the circumstances of that specific situation.

Following the DPA 2018's designation of universities as public authorities, this stipulation in Recital 43 has been taken seriously. Guidance issued by the Information Commissioner's Office[103] suggests that public interest may be the most appropriate ground for processing personal data for academic research. This recommendation has in turn been reflected in guidance issued to researchers by a number of UK universities.[104]

While the requirement that consent be 'freely given' is not unique to public authorities, any universities classified as public authorities for the purposes of the GDPR (of which there will be many across Europe and beyond) are particularly encouraged to give thought to whether their 'balance' of power with the data subject means they can appropriately rely on this basis in practice.[105] This encouragement to public authorities cannot be dismissed as non-binding exhortation, as authoritative guidance has been issued by the Article 29 Working Party, and subsequently by UK regulators, to the effect that public authorities should avoid reliance on consent as a basis for processing in the performance of their tasks.

Whether it is truly impossible for consent to be freely given in the context of academic research requires careful consideration. Guidelines issued by Article 29 Working Party stress that it is essential that individuals who refuse consent must not be denied access to the public authority's core services.[106] For consent to be '*freely given*' it is essential that individuals who refuse to consent are not denied access to the public authority's '*core services*.' This risk of detriment is easily identifiable in the relationship between students and their

98 GDPR, art 6(3).

99 Arts 37–38

100 Arts 13 and 14(1)(b).

101 Recital 145 and art 83(7); see Butler (n 2).

102 Ibid, art 4(11).

103 Information Commissioner's Office, 'Guide to the General Data Protection Regulation (GDPR)' (September 2018) <https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf> accessed 2 October 2018.

104 For example, University College London, 'Guidance for Researchers on the Implications of the General Data Protection Regulation and the Data Protection Act 2018' <https://www.ucl.ac.uk/legal-services/sites/legal-services/files/guidance_paper_for_researchers_0.pdf> accessed 6 November 2018.

105 GDPR, Recital 43.

106 Art 29 Working Party, 'Guidelines on consent under Regulation 2016/679' (WP 259, 10 April 2018), at 6.

institution, but is less obvious in the context of academic research. Naturally, the risk of detriment could vary depending on the nature of the research. An individual participating in a clinical trial in the hope of combatting life-threatening illness clearly has a much greater dependence on a 'core service' than an interviewee in a qualitative study who stands to gain no personal benefit from the research.[107]

Admittedly, concerns about appropriate power balance are not the only reason why UK researchers have been steered away from consent as a basis for processing. The GDPR also requires granularity of consent, and it has been suggested that not every research project is sufficiently defined at the point when consent would be collected from participants to fulfil the enhanced requirements of the GDPR.[108] The requirement to delete data should consent be withdrawn could also prove problematic in cases where data are required to be retained post publication. Nevertheless, should researchers wish to process participant data on the basis of consent, their balance of power with the data subjects will be a very relevant consideration under the GDPR.

Clearly, the role of consent within academic research needs to be re-evaluated if it is not widely commended as a legal basis for processing personal data. Consent may still be required under the common law duty of confidentiality, even when it is not a GDPR basis for processing. This could help to highlight to researchers the *purpose* of obtaining consent from participants as a justification for using private information, as opposed to a bureaucratic exercise. Reconsideration of the role of consent could also take into account empirical studies of participant views as to why they value consent: as a means for securing their approval of research projects;[109] as a mechanism for respecting their autonomy;[110] or even simply as an act of courtesy during the research project.[111]

Regardless of how great a role consent will continue to play in academic research, however, the above analysis illustrates the importance the new obligations placed upon public authorities by the GDPR in an academic context. As the threshold of the exemptions made under the relevant derogations is set high, particularly where private information is used, these obligations are therefore likely to be of greater import to academic research in practice than any new freedoms the GDPR may bring.

Conclusion

In sum, it appears that the GDPR derogations which could, hypothetically, provide for greater freedom in academic research data processing in fact set the bar high for their exercise by Member States. The Regulation only allows for derogation to the extent that it is 'necessary' for academic expression, or would cause impossibility or serious impairment to research. We have seen from the UK's example that these restrictions in the GDPR, combined with the impact of privacy rights wherever they may be engaged by research, significantly limit the scope of academic freedom in personal data processing.

The answer to the question posed at the beginning of this paper—'do academic researchers stand to have more freedom in their data processing under the GDPR?'—therefore appears to be 'no', at least as far as can be established from analysis of the UK's implementation of the Regulation. The derogations should rightly be seen as the exceptions which prove the rule of public authority obligations which apply to many universities within the territorial scope of the Regulation. Our secondary lesson is therefore that the innovations in the GDPR of more significance for academic research are the provisions specifically relating to public authorities. These shifts may be subtle in their impact, but the need for researchers to consider appropriate balances of power with data subjects cannot be ignored under the new Regulation.

doi:10.1093/idpl/ipz010
Advance Access Publication 15 July 2019

107 Although, as the GDPR acknowledges, in the case of clinical trials consent would be governed by Regulation (EU) No 536/2014.

108 Leslie Stevens, 'The Proposed Data Protection Regulation and Its Potential Impact on Social Sciences Research in the UK' (2015) 1 European Data Protection Law Review 97; David Erdos, 'Systematically Handicapped? Social Research in the Data Protection Framework' (2011) 20 Information and Communications Technology Law 83.

109 Mhairi Aitken and others, 'Public Responses to the Sharing and Linkage of Health Data for Research Purposes: A Systematic Review and Thematic Synthesis of Qualitative Studies' (2016) 17 BMC Medical Ethics 73.

110 Amy L McGuire and others, 'DNA Data Sharing: Research Participants' Perspectives' (2008) 10 Genetics in Medicine 46.

111 Gill Haddow and others, '"Nothing Is Really Safe": A Focus Group Study on the Processes of Anonymizing and Sharing of Health Data for Research Purposes' (2011) 17 Journal of Evaluation in Clinical Practice 1140; Michael Robling, 'Public Attitudes towards the Use of Primary Care Patient Record Data in Medical Research without Consent: A Qualitative Study' (2004) 30 Journal of Medical Ethics 104.

Publication 6

Publication Title: Secondary Uses of Patients' Data in the European Health Data Space: A UK-German Comparison

Short Title: 'EHDS Chapter'

Page Number of Thesis: 76

The following pages incorporate a PDF of this book chapter. The edited collection in which it was published, *Confidentiality, Privacy, and Data Protection in Biomedicine: International Concepts and Issues* (Dove, ed), has not been made public on an open-access basis. This publication will therefore be redacted in the published version of this PhD thesis.

The formatting, pagination and footnote numbering of the publication have been retained, and thus the numbering is self-contained. The pagination sequence of this thesis will then resume at the end of the publication.

Publication 7

Publication Title: The ‘DPIA+’: Aligning Data Protection with Equality Law

Short Title: ‘DPIA+’ Article

Page Number of Thesis: 77

The following pages incorporate a PDF of a work-in-progress version of this article. This version will be edited following peer review, which recommended minor revisions. I plan to resubmit this article to the *Computer Law & Security Review* in August or September 2025.

The formatting, pagination and footnote numbering of the publication have been retained, and thus the numbering is self-contained. The pagination sequence of this thesis will then resume at the end of the publication.

The 'DPIA+': aligning data protection with equality law

Abstract

Data protection is often characterised as focused on individuals' rights. This individualistic account has prompted concerns that laws such as the General Data Protection Regulation ('GDPR') are inadequate to regulate the more systemic consequences of data processing, such as data-driven discrimination on the basis of race, disability or gender. This critique can miss opportunities to emphasise and develop the obligations which the GDPR imposes, such as the Data Protection Impact Assessment ('DPIA').

This paper presents the case for the DPIA as a key evaluative tool, particularly within the public sector. Using the UK as a case study, I show how the DPIA's minimum requirements under Article 35 GDPR can be enhanced through aligned consideration with equality law. This alignment allows risks of discrimination stemming from data bias to be considered holistically, through a combined 'DPIA+', before public authorities begin a significant new programme of data processing. By way of example, I review NHS England's Covid-19 data store, which ran from 2020 to 2022. The focal case study of the paper thus revolves around the health service, but the wider argument for a shift in evaluative practice has wider implications across the public sector.

Key Words: Big Data, Data Protection, Data Protection Impact Assessment, Discrimination, Equality Law

1. Introduction

The last decade has seen broad, multidisciplinary discussion of 'Big Data': the algorithmic analysis of large volumes of information about people, with the promise of yielding faster (even, perhaps, better) insights.[1] An argument often made within this discourse is that privacy and data protection instruments are too 'small', too granular,[2] to regulate Big Data. Broadly speaking, these laws are often seen as typified by data subjects' rights, and thus centred around the interests and choices of the individuals identified by the data. As such, the general argument goes, data protection law atomises regulation into consideration of the risks and harms to each individual, who may or may not have the information, insight and expertise to regulate the use of their own data.[3] A common corresponding claim is that data protection therefore provides inadequate protection for the interests of groups,[4] for individuals subject to downstream consequences of data processing,[5] or for society as a whole. A range of supplements/alternatives have been proposed to remedy this alleged deficiency, from

[1] 'Big Data' is often characterised according to volume, variety, and velocity—some authors have also added 'value' and 'variability'. There is a broad agreement that the data generated by healthcare services (the case study for this paper) contains these features, see e.g. Blagoj Ristevski and Ming Chen, 'Big Data Analytics in Medicine and Healthcare' (2018) 15 Journal Integrative Bioinformatics 3; Pantea Keikhosroki (ed). *Big Data Analytics for Healthcare : Datasets, Techniques, Life Cycles, Management, and Applications* (2022) (London, Academic Press).

[2] Julie E Cohen, 'Turning Privacy Inside Out' (2019) 20 Theoretical Inquiries in Law 1, 1-22.

[3] Michael Froomkin, 'Big Data: Destroyer of Informed Consent' (2019) 18 Yale Journal of Health, Policy, Law and Ethics 3. See also notes 4-9 below.

[4] Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (2017) (Cham, Springer).

[5] Aisling McMahon, Alena Buyx, and Barbara Prainsack, 'Big Data Governance Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond' (2020) 28 Medical law review 1, 155–182; Megan Doerr and Sandra Meeder, 'Big Health Data Research and Group Harm: The Scope of IRB Review' (2022) 44 Ethics and Human Research 4, 34-38.

additional data subject rights to help individuals control algorithmic inferences about them,[6] to algorithmic impact assessments,[7] and governance models such as data trusts[8] or systemic oversight.[9]

This paper challenges the individualistic account of data protection law, at least to the extent that it applies to the EU and UK General Data Protection Regulation.[10] Focusing on the UK to permit more detailed analysis of a case study, I argue that the Data Protection Impact Assessment[11] can be an important tool to identify bias and harm at a group level. This argument is made in the context of (some of) the 'Biggest' Data in the UK: the National Health Service ('NHS') datasets for patients in England. I focus on the NHS Covid-19 'data store', which is now being migrated into a longer-term resource. The focus on this case study allows for more detailed consideration of the risks of data processing. However, the implications of the argument go beyond the health sector, as any public authority will need to consider the impact of their large-scale data processing.[12]

The Covid-19 Data Store was urgently set up to help the UK government decide how to distribute resources at the beginning of the pandemic.[13] We know, now, that some groups fared worse in the pandemic, and had disproportionately high rates of mortality.[14] There was clearly a risk—and to be clear, I only propose it as a *risk*—that bias in the data could have misrepresented the prevalence of Covid-19 in some populations, and thus misrepresent their need for (for example) more clinical staff in their area, or prioritisation for vaccination.[15] This risk has been explored, retrospectively, in epidemiological studies. But the crucial point for this paper is that the GDPR requires that accuracy, fairness and risks of discrimination be considered *prospectively*, particularly within a Data Protection Impact Assessment ('DPIA').[16] A DPIA was conducted by NHS England for the Covid-19 data store, but the published document does not evidence any reflection on risks associated with the quality of the data.

The central argument of this paper is, therefore, that when public authorities use Big Data to shape policy, and make significant decisions about people,[17] the risk of bias and downstream (indirect) discrimination should be investigated through a combined lens of data protection and equality law. By itself, data protection is already an effective tool by requiring a DPIA which should cover the

---

[6] Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review, 494-619.
[7] Ada Lovelace Institute, 'Algorithmic impact assessment: user guide' (8 February 2022) available from: https://www.adalovelaceinstitute.org/resource/aia-user-guide/ (accessed 12 March 2025).
[8] Sylvie Delacroix and Neil D Lawrence, 'Bottom-up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance' (2019) 9 International Data Privacy Law 4, 236–252.
[9] Effy Vayena and Alessandro Blasimme, 'Health Research with Big Data: Time for Systemic Oversight' (2018) 46 The Journal of Law, Medicine & Ethics 1, 119-129.
[10] 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 (United Kingdom General Data Protection Regulation); hereafter cited as 'the GDPR.'
[11] Ibid, Article 35.
[12] Under Article 35 GDPR, a data protection impact assessment is mandatory when any data controller processes 'special categories' of personal data on a large scale. As these special categories include personal data revealing racial or ethnic origin, as well as data relating to health, much public authority processing activity will be captured by this requirement (particularly if information, e.g. names, are used as a proxy for ethnic background).
[13] Matthew Gould, Indra Joshi and Ming Tang, 'The power of data in a pandemic'. (28 March 2020) available from: https://digileaders.com/the-power-of-data-in-a-pandemic/ (accessed 12 March 2025).
[14] Public Health England, 'Disparities in the risk and outcomes of COVID-19' (August 2020): Disparities in the risk and outcomes of COVID-19 (publishing.service.gov.uk) (accessed 12 March 2025).
[15] Elizabeth Williamson et al, 'Risks of covid-19 hospital admission and death for people with learning disability: population based cohort study using the OpenSAFELY platform' (2021) 374 BMJ 1592.
[16] Article 35 GDPR, discussed in detail in section 3.
[17] For ease of reference, I am taking this phrase from Article 22 GDPR, which encompasses decisions producing legal or other significant effects.

downstream consequences of the processing for society as a whole.[18] But it could be strengthened by combining it with obligations under equality law— in the UK, this means consideration of the Public Sector Equality Duty (or 'PSED') under s.149 Equality Act 2010.

The overlap between these frameworks is not a fabrication of my own conjecture, and has already occurred in the case of in *R (Bridges) -v- CC South Wales & ors*, in which the legality of police use of automated facial recognition technologies was challenged.[19] This was not a case from the healthcare sector, but the same legal obligations applied to South Wales Police as are discussed here in the context of the NHS. The Court of Appeal's judgment in this case illustrates how failure to assess the risk of discriminatory bias in software had parallel implications in data protection and equality law. It indicates that an inadequate DPIA has been conducted under the GDPR (as was the case for South Wales Police[20]). At the same time, this lack of investigation into the risk of algorithmic discrimination can also be a failure to comply with the PSED— as was held to be the case in this instance.[21]

Although, on the surface, the use of Big Data across the public sector is varied, in reality the same risks associated with under-representation, or over-representation, in a dataset arise. The same groups of people who might be over-flagged by automated facial recognition (per *Bridges*) might also be under-represented in the analytics informing Covid-19 protective measures. These risks are complex, and require careful, expert evaluation. It is therefore worth exploring the mutual reinforcement of DPIA and PSED requirements, which both touch on risks of discrimination. To avoid the acronym soup of DPIA-PSED, especially when the additional letters of Equality Impact Assessments[22] and Al Impact Assessments[23] could also become incorporated, I summarise the net effect of their alignment as 'DPIA+.'

The next section explains in more detail why the risk of bias in public-sector data warrants the enhanced safeguard of a DPIA+. The resulting disparities in public services can have life and death implications for the individuals involved.

2. Background: New Models of Evaluation

2.1 The Risks of Public-Sector Automation

To explain why I (along with others, see 2.2 below) have identified a need for new tools to evaluate Big Data processing, this section will give a brief outline of the risks of bias and discrimination in public-sector information.

The risks of rolling out software to evaluate people, and allocate public benefits, across a whole national population are significant. A vivid example comes from the Dutch welfare system, which introduced algorithms that inaccurately labelled dual national and/or 'foreign-sounding' recipients as likely benefit frauds. The fault in question can be characterised as one of bias, as the software used

[18] According to the UK's data protection regulator—the Information Commissioner's Office (see below, note 51).
[19] [2020] EWCA Civ 1058.
[20] The Court made this finding on the ground presented in submissions: that the DPIA was inadequate due to its erroneous assumption that Article 8 European Convention on Human Rights was not engaged. It is possible that the Court would also have found the DPIA lacking due to its failure to assess the risk of harm to data subjects in the form of discrimination, had that argument been presented to them.
[21] Note 19, [202].
[22] A non-mandatory form of report in which compliance with the public sector equality duty can be demonstrated via written record of the deliberations—see below note 37.
[23] Not a statutory term, but another form of impact assessment proposed by, for example, the UK Government's Office for Artificial Intelligence, 'Guidelines for AI procurement' (8 June 2020) available from: https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement (accessed 12 March 2025).

ethnicity as an (illegal) proxy for dishonesty when making its assessments.[24] As with the UK's own Post Office scandal,[25] the victims of these automated errors reported lives ruined, and even suicide.[26] It is entirely possible for comparable bias to occur in other public authorities: the UK's Department for Work and Pensions has also reportedly used an AI to detect benefit fraud, and subsequently undertaken a fairness analysis that showed up bias on the basis of age, disability, marital status and nationality.[27]

There are, undoubtedly, advantages to automated analysis within public services—for example, by detecting disease early or improving patient screening programmes.[28] But just as it is true to say that 'data saves lives,'[29] inaccurate and discriminatory use of data can evidently ruin them, and even cut them short. There is no evidence that the algorithms used by UK public authorities so far[30] have caused harm to an equivalent degree as the Dutch welfare automation. But, clearly, some life-and-death processes are entrusted to automated systems. As an example, the Information Commissioner's Office (the data protection regulator for the UK) noted that an inadequately-tested code was released into the NHS transplant system and failed to match five patients before it was caught.[31] Again, fortunately, these five patients were reported to not to have suffered any harm, but it is reasonable to infer that any flaws in software applied nationally could have had much wider consequences.

2.2 New Assessment Models

The above is just a brief overview of the individual and societal harms which can be caused through the (mis)application of Big Data analytics across a national population. The risks of data-driven services in the public sector, and beyond, have been much discussed. Consequently, several authors[32] have already argued for expanding the scope of the Data Protection Impact Assessment ('DPIA') under the GDPR. I agree with Mantelero that the concerns captured within DPIA's should be broadened to capture the consequences of data processing, and not just its immediate quality and security.[33] I also agree with Bertaina and colleagues, that the forthcoming Fundamental Rights

[24] European Parliament, Parliamentary question O-000028/2022, 'The Dutch childcare benefit scandal, institutional racism and algorithms' (28 June 2022) available from: https://www.europarl.europa.eu/doceo/document/O-9-2022-000028_EN.html (accessed 12 March 2025).
[25] Jane Croft, 'Post Office tried to 'hush up' case of worker who killed himself, inquiry hears' (26 April 2024), available from: https://www.theguardian.com/uk-news/2024/apr/26/post-office-tried-to-hush-up-martin-griffiths-case-inquiry-hears (accessed 12 March 2025).
[26] Melissa Heikkilä, 'Dutch scandal serves as a warning for Europe over risks of using algorithms' (29 March 2022) available from: https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/ (accessed 12 March 2025).
[27] Robert Booth, 'Revealed: bias found in AI system used to detect UK benefits fraud' (6 December 2024) available from: https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits (accessed 12 March 2025).
[28] NHS England, 'About the NHS AI Lab' https://transform.england.nhs.uk/ai-lab/about-the-nhs-ai-lab/ (accessed 12 March 2025).
[29] Department for Health and Social Care, 'Data saves lives: reshaping health and social care with data' (15 June 2022) available from: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data (accessed 12 March 2025).
[30] The Public Law Project has obtained evidence of 55 automated systems used by the UK government, see 'The Tracking Automated Government register' (updated October 2023) available from: https://publiclawproject.org.uk/resources/the-tracking-automated-government-register/ (accessed 12 March 2025).
[31] Information Commissioner's Office, 'NHS Blood and Transplant' (3 March 2023) available from: https://ico.org.uk/action-weve-taken/enforcement/nhs-blood-and-transplant/ (accessed 12 March 2025).
[32] Alessadro Ortalda and Paul De Hert, 'Artificial intelligence human rights impact assessment' in Alberto Quintavalla and Jeroen Temperman (eds.), *Artificial intelligence and human rights* (2023) (Oxford, Oxford University Press). See also notes 33 and 34 below.
[33] Alessandro Mantelero, 'AI and Big Data: A blueprint for a human rights, social and ethical impact assessment' (2018) 34 Computer Law & Security Review 4, 754-772.

Assessment under the EU's AI Act offers a promising supplement to a DPIA in the context of AI,[34] especially as it concentrates attention on groups who may be particularly affected by the processing.[35] The AI Act, will not, naturally, apply within the UK.

The main, distinguishing contribution of this article is its proposed use of equality law as a supplement to the DPIA, as opposed to fundamental rights in general. I suggest equality law as a vehicle for these fundamental rights because, at least in the public sector, failure to consider the impact on groups is not just an issue for a DPIA, but also a potential breach of obligations under equality legislation. The alignment between these two areas of law has been under-explored, and equality law can bring a wealth of existing guidance to those wishing to expand on the minimum requirements the DPIA, as set out in data protection law.

The evolution of equality law across Europe is complex and varied, but nonetheless shares much common origin. Of the pre-Brexit EU/ EEA countries, only the UK, Denmark and Norway lacked constitutional provisions regarding equality and non-discrimination.[36] Even in countries without these explicit constitutional principles, the prohibition on discrimination under Article 14 European Convention on Human Rights ('ECHR') will form a fundamental part of the national law. In the UK, the key statute is the Equality Act 2010, which sets out the 'public sector equality duty' at s.149. This duty has many parallels to the DPIA within the public sector. It requires proactive consideration of the impact of a new measure or policy on different groups protected under the Equality Act (on grounds of race, gender, disability etc.). It has prompted the development of voluntary 'Equality Impact Assessments,'[37] which (like a DPIA) provide both a structure for an evaluation of impact, and also a means of recording and demonstrating this evaluation for the sake of accountability.[38]

The advantage of focusing on one country, as this paper does, is that only this version of equality law needs to be considered, without the potential distraction of evaluating the significance of any national variation. It also means that a single, factual case study can be explored in greater detail. In this regard, I concur with Bertaina and colleagues that consideration of new assessment models in light of concrete examples is important to advance a field of study in relative infancy.[39] That said, however, the common origins of anti-discrimination principles under EU and ECHR law make it likely that alignment between data protection and equality law would have similar impact in other jurisdictions. This could be studied further in future work.

The following analysis is, therefore, an initial exploration of the value of aligning DPIA's with equality law in the public sector. I am not advocating a particular form or process for this combined evaluation: partly, because this is beyond the scope of this article. Also, because the public sector is broad, and the risks of discrimination may differ significantly according to context. It is therefore beneficial to have a degree of flexibility as to implementation. Furthermore, where public authorities are already conducting equality impact assessments (or their equivalent outside the UK), I would hope

[34] Samuele Bertaina, Ilaria Biganzoli, Rachele Desiante, et al, 'Fundamental rights and artificial intelligence impact assessment: A new quantitative methodology in the upcoming era of AI Act' (2025) 56 Computer Law & Security Review 106101.
[35] Under Article 27.1(c) of the AI Act. See note 34 above.
[36] Christopher McCrudden and Sacha Prechal, 'The Concepts of Equality and Non-Discrimination in Europe: A practical approach' (2009) available from: https://ec.europa.eu/social/BlobServlet?docId=4553 (accessed 12 March 2025), 3.
[37] Doug Pyper, 'The Public Sector Equality Duty and Equality Impact Assessments' (2020) available from: https://researchbriefings.files.parliament.uk/documents/SN06591/SN06591.pdf (accessed 12 March 2025).
[38] 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' WP 248 (Brussels, 4 April 2017).
[39] Note 34.

that these existing processes could be combined with a DPIA, so that a 'DPIA+' can represent a consolidation of effort, rather than additional 'red tape.'

2.3 Automation & Covid-19

The case study explored in this paper is the UK's NHS Covid-19 data store. This resource is not explored to make any suggestion of discrimination in the NHS response to Covid-19. Rather, the risks of under-representation of some demographics that have since become apparent are highlighted, to establish a prima facie case for a risk of bias, worthy of evaluation to prevent discrimination used.

The UK government announced in March 2020 that it would create an NHS Covid-19 'Data Store' from information routinely collected as part of the health service. This 'Store' would use data from the NHS (and other public authorities) to create 'dashboards' with predictions developed through artificial intelligence. These dashboards would in turn provide a '*single source of truth*'[40] about the spread of the coronavirus in England. Decisions made on the basis of these dashboards would be significant, even (it was suggested) to the point of diverting patients and critical resources between hospitals based on their predictions.

The data store was 'Big Data' in the classic sense of aggregating a variety of data to build a bigger picture of a national trend, which would be updated daily with fresh information.[41] Machine-learning tools were built into the platform using Microsoft Azure to make predictions about the spread of the virus. By going beyond a factual picture of the outbreak, and attempting to anticipate it using AI, the government was, in essence, conducting fast-paced, real-time epidemiology on the spread of Covid-19. The tone of the announcement of the store conveys the urgency of the undertaking, and the vital resources at stake:

The NHS is facing an unprecedented challenge. Responding to the Covid-19 crisis will require everything we have and more. In the fight against this pandemic, decision-makers will need accurate real-time information. To understand and anticipate demand on health and care services, we need a robust operating picture of the virus, how it's spreading, ***where it might spread next*** *and how that will affect the NHS and social care services. On the supply side, we need to know* ***where the system is likely to face strain*** *first, be that on ventilators, beds or staff sickness* (emphasis added).[42]

We now know that different patient groups experienced different outcomes in the first wave of Covid-19. It is not the aim of this article to raise any causative links between these outcomes and the allocation of resources. Rather, these differences are explored in order to make a case that the risk of disparity could have been anticipated, and usefully explored in a DPIA (as enhanced by equality law). To explain why, however, the next section provides more information on the nature and purpose of the DPIA.

3. The Data Protection Impact Assessment

3.1 DPIA's and 'Downstream' Consequences

This paper concurs with other authors[43] that the DPIA can be understood expansively, as concerned with the wider consequences of data processing, and not just its immediate quality and security. In order to accept this position, it is necessary to understand the GDPR as an instrument which is not entirely focused on individuals, and can also regulate equal treatment between groups. While individual data subjects are important, a large proportion of the GDPR is devoted to the obligations of the organisations which control data: 'data controllers.'[44] The rights of individuals identified by the

[40] Note 13.
[41] See the general definition of Big Data, note 1.
[42] Note 28.
[43] Particularly Mantelero, note 33.
[44] Article 4(7), GDPR.

information—'data subjects'— in fact make up only 9 of its total 99 Articles.[45] Delacroix and Lawrence suggest the resulting data controller-data subject relationship is 'feudal,'[46] but this balance of provisions does not need to be understood in this way. As these authors themselves acknowledge,[47] the individual cannot be fairly or reasonably expected to regulate all of the data processing of which their information may form a part.

Rather, it is the data controller—not any of the individuals identified by the information—who has the necessary overview of all the information they process, the time and resources to evaluate associated risks to individuals and groups, and most importantly, the legal responsibility to ensure their lawful use of information. Individuals have some rights to modify how information about them—and only them—is used, as a recognition of their personal right to privacy and data protection.[48] But, as the UK Supreme Court has pointed out, this does not mean that data protection law as a whole is predicated on the idea of individual control.[49]

Ultimately, data controllers have the responsibility to ensure that their use of personal information is lawful. They must make this determination with reference to the GDPR's broad principles of transparency, lawfulness, data minimisation, accountability, and (crucially for the anti-discrimination questions raised in this paper) fairness.[50] These principles, I suggest, go beyond the interests of the individual and address the more widespread implications of data processing.

As a key example, data protection impact assessments are not limited to, or even primarily focused on, the risks to a collection of individuals. As the UK's Information Commissioner's Office ('ICO') notes in their guidance:

It is also important that you don't just consider obvious and immediate tangible damage to people. But also more subtle intangible harms and how the system might affect people's rights and freedoms more generally. This includes any impact on society as a whole. For example, DPIAs require you to consider risks to rights and freedoms of all those that the system might affect.[51]

This guidance provides an important gloss on the text of the GDPR. By opening up the scope of consideration beyond the rights of the individuals identified by the information processed, the ICO is also making an important point about the broader purpose of data protection law. Data subjects are clearly important, as the people most immediately impacted when information about them is used. But they are not the only people impacted by the potential downstream consequences of—for example—a policy decision made using some people's personal data, but then applied much more widely. Hence why McMahon and colleagues advocate for new bodies to deal with the downstream consequences of Big Data.[52]

I am not dismissing the question of further regulation for these kinds of downstream consequences. My argument, however, is that existing data protection legislation—and, in particular, the DPIA—contains untapped potential to address consequences of processing which go beyond the individuals

[45] Articles 12-23 GDPR.
[46] Note 8, 239.
[47] Ibid.
[48] Under the European Convention on Human Rights, and the European Union Charter of Fundamental Rights, as acknowledged in the preamble to the GDPR.
[49] *Lloyd v Google* [2021] UKSC 50, [108-109]. This judgment was decided on the Data Protection Act 1998, which was based on the previous EU Data Protection Directive, but the GDPR has not significantly altered this fundamental balance of responsibility.
[50] Article 5 GDPR.
[51] ICO Guidance, 'How do we ensure lawfulness in AI?' (28 October 2024) available from: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-lawfulness-in-ai/ (accessed 12 March 2024).
[52] McMahon et al, note 5.

identified by the information. The ICO's guidance provides a helpful nudge to data controllers to consider the 'bigger picture' of their processing—such as the risk of data bias skewing their downstream policy and decision-making. This nudge deserves highlighting, amplification, and (crucially), formalisation, so that these wider considerations are considered systematically. The next subsection explains further why the DPIA is a helpful nexus for this formalisation.

3.2 The Data Protection Impact Assessment

The Data Protection Impact Assessment was introduced by the EU's General Data Protection Regulation in 2018. As such, it is a relatively new instrument, albeit it was preceded by a (non-mandatory) Privacy Impact Assessment under the previous Data Protection Directive.

It is not misleading to describe the DPIA as an 'obligation,' as the 'Big Data' scenarios of large-scale processing discussed in this paper will trigger the DPIA as a mandatory requirement under Article 35 GDPR. A data controller must undertake a DPIA to before they embark on large-scale processing of 'special category'[53] data. If a high risk to data subjects is identified, the controller can only proceed subject to the instructions of the national data protection regulator.[54] The DPIA process is clearly intended by the drafters of the GDPR to be a serious gatekeeping exercise for high-risk data processing, with much sharper 'teeth' than mere good-practice guidance. The mandatory nature of these legal requirements is a key to their value as safeguards— providing a reassuring contrast to the voluntary measures for developers the UK Government has otherwise proposed in the context of the development of AI.[55]

A data controller must conduct a DPIA when they process a large volume of 'special category' personal data—this will include data relating to people's health, as well as their race, ethnicity or sexual orientation.[56] To summarise the requirements of a DPIA under Article 35(7) of the GDPR, the assessment must contain at least:

> *a) a systematic description of the envisaged processing*
>
> *b) an assessment of the necessity and proportionality of the processing*
>
> ***c)** an assessment of the risks to the rights and freedoms of data subjects **referred to in paragraph 1***
>
> *d) the measures envisaged to address the risks* (emphasis added).

For the sake of this paper it is subparagraph (c) which is particularly significant: the 'assessment of risks and freedoms to data subjects.' 'Data subjects,' here, does not appear to refer only to the 'immediate' data subjects, whose information will be used for the processing; it also seems to leave the door open for downstream data subjects. The data subjects referred to in paragraph 1, which only specifies '*a high risk to the rights and freedoms of **natural persons***' (emphasis added). As this paragraph does *not* confine consideration of risk to the people whose information is immediately at stake, it seems entirely reasonable to infer, as the ICO advises, that it encompasses risk to people across society, including those caught by downstream consequences of the processing. As discussed above, an example would be when the data processing is used to generate a prediction about a future

[53] Often called 'sensitive data' in practice, this term under Article 9 GDPR covers health-related data, as well as information about sex life, sexual orientation, race or ethnicity, all of which are likely to be included (to some extent) in an individual's healthcare record.
[54] Article 36, GDPR.
[55] Although some update has been achieved through the introduction of the AI Safety Institute. See Department for Science, Technology and Innovation 'AI Opportunities Action Plan' (13 January 2025) available from: https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan (accessed 12 March 2025).
[56] Article 9 GDPR.

trend in the population—such as the spread of a disease—and the 'downstream' consequences include who does, and does not, receive additional resources in their local hospitals.[57]

3.3 The DPIA & Fairness

As a GDPR obligation, a DPIA must be conducted in a way that seeks alignment with the principles of the regulation. Of the overarching principles set out in Article 5 GDPR, the fairness principle[58] is particularly important for our current purposes. It is a principle which some place at the heart of data protection law, as a distinct body of doctrine.

For example, in seeking to define the essential core of the right to data protection, as distinct from privacy, Tzanou lands on 'fair processing' as a starting point, and also includes the principle of non-discrimination.[59] Dove broadly concurs with this characterisation of data protection, as distinct from privacy.[60] Lynskey, in considering the 'added-value' of the right to data protection, suggests that there are some harms that data protection wards against more effectively, such as the risk of discrimination through proxies or presumptions.[61]

I agree with these authors that the distinction between privacy and data protection is complex, and difficult to pin down, but that there is still a non-symbolic difference between the two.[62] A final pronouncement on this question is beyond the scope of this paper, but one key aspect of the distinct character of data protection law lies in its attempt to achieve fairness between different actors and stakeholders. Data subjects are given individual rights,[63] data controllers a carefully codified framework through which to process data lawfully,[64] and even national regulators have their roles and powers set out.[65] As the scope of the GDPR goes well beyond the 'micro' dynamics of an individual controller and subject, it is less surprising that its core principles include more systematic considerations, such as fairness and accountability.[66]

The GDPR's conception of fairness seems to be, at the very least, congruent with that in equality law, and it seems from the text of the GDPR that the two overlap. Within the recitals, it is explained that 'special categories' of information (the large-scale use of which prompts a DPIA[67]) because their use gives rise to a higher risk of discrimination.[68] Recital 71 of the GDPR contains a long, heavily caveated sentence which truncates to:

In order to ensure fair and transparent processing in respect of the data subject, […] *the controller should* […] *secure personal data in a manner* […] *that prevents, inter alia, discriminatory effects*

Additionally, Recital 75 GDPR makes it clear that the risk of discrimination is intended to be considered among the harms the controller should consider as part of a DPIA:

---

[57] As appeared to the case with the NHS Covid-19 Data Store, which is considered in section 4.
[58] Or, at least, the combined principle of lawfulness, fairness and transparency under Article 5 (1)(a).
[59] Maria Tzanou, 'Data protection as a fundamental right next to privacy? "Reconstructing" a not so new right' (2013) 3 International Data Privacy Law 2, 88–99, 89.
[60] Edward S Dove, 'The EU General Data Protection Regulation: implications for international scientific research in the digital era' (2019) 46 Journal of Law, Medicine & Ethics 4, 1013-1030.
[61] Orla Lynskey, 'Deconstructing data protection: the "added-value" of a right to data protection in the EU legal order' (2014) 63 International and Comparative Law Quarterly 3, 569-597.
[62] Juliane Kokott and Christoph Sobotta, 'The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 4, 222–228.
[63] Articles 12-22 GDPR.
[64] Articles 24-43 GDPR.
[65] Articles 51-59 GDPR.
[66] Article 5 GDPR.
[67] Article 35.3(b) GDPR.
[68] Recital 51 GDPR.

The 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

Binns has characterised the DPIA as a form of 'meta-regulation': an attempt by the state to shape corporations' internal efforts to assess and self-regulate their use of data.[69] This may well be a useful characterisation of the DPIA in its private-sector manifestation, but does not capture all the dimensions of a public-sector DPIA. Similar to a private company, a public authority will bear its own legal responsibility to comply with their responsibilities as a data controller, and is encouraged to take ICO guidance into account.[70] However, public authorities will have an additional level of scrutiny in the shape of judicial review of their decisions via the High Court in England and Wales where— as emphasised in this paper— other public law considerations can be brought to bear. It is perfectly possible, for example, for a Claimant to plead a breach of the GDPR and the PSED in relation to the same planning process around a new policy.

The significance of the DPIA should not come as a surprise to anyone with a familiarity with data protection as it has been applied in the public sector. It is a record of review of the legality of data processing, which can be considered by a court in the event of judicial review into a new policy.[71] Failure to conduct an adequate DPIA is a breach of the GDPR, rendering processing unlawful even if it is based on government policy.[72] The importance of the DPIA when new public policy involves large-scale processing of personal data should not be underestimated, as a superficial, unreflective 'tick-box' review of risks to data subjects is unlikely to pass muster within a judicial review. This is illustrated further in the next section.

4. The PSED and the DPIA+

4.1 The PSED

The Public Sector Equality Duty was introduced in the Equality Act 2010. It contains some of the lineage of previous non-discrimination duties in English law.

Like the DPIA obligation, the PSED is essentially a reflective requirement: a necessary pause before an action is taken, to ensure appropriate evidence of its lawfulness. Before the relevant action is taken, section 149 of the Equality Act 2010 requires public authorities to have 'due regard' to their duty to:

(a)eliminate discrimination, harassment, victimisation and any other conduct that is prohibited by or under this Act;

(b)advance equality of opportunity between persons who share a relevant protected characteristic and persons who do not share it;

(c)foster good relations between persons who share a relevant protected characteristic and persons who do not share it.

The duty applies when a public authority 'exercises its functions.' In practice, this covers most of the authority's outward-facing activity concerning the public, as opposed to its internal functions (e.g. as

[69] Reuben Binns, 'Data protection impact assessments: a meta-regulatory approach' (2017) 7 International Data Privacy Law 1, 22–35.
[70] The ICO are, strictly speaking, a regulatory authority independent of the government, but they nonetheless fall within an expansive understanding of the state.
[71] For example, when the expansion of the Ultra Low Emission Zone was judicially reviewed in *London Borough of Hillingdon & Ors, R (On the Application Of) v Mayor of London (Re ULEZ Expansion)* [2023] EWHC 1972 (Admin).
[72] *R(HM) v Secretary of State for the Home Department* [2022] EWHC 695 (admin).

an employer). When a public authority processes a large volume of personal data for analytical purposes, this will almost certainly be in the exercise of its functions. As such, it is safe to assume that the use of Big Data and AI in the public sector will be subject to the PSED.

4.2 Bridges v South Wales Police

To illustrate the DPIA-PSED alignment in practice, it is helpful to look outside the healthcare sector, to a case where both obligations have been challenged within the Courts. In this example, South Wales Police (also a public authority) were judicially reviewed for their failure to conduct an adequate DPIA, and for breaching the PSED. A review of this case suggests that these two legal breaches arose, in fact, from the same omission: a failure to systematically review the risks of the software they used on the general public.

A key strength of the DPIA is its capacity to prompt reflection when novel technologies are deployed by public authorities. An inadequate DPIA thus represents a missed opportunity to review the evidence available, before data processing begins. The High Court judgment in *Bridges v South Wales Police* was suggested to be the first time any court in the world had considered the use of Automated Facial Recognition ('AFR') Software.[73] The claim for judicial review of South Wales Police's use of AFR was brought by Edward Bridges, with the support of the human-rights non-governmental organisation Liberty.

The subsequent judgment of the Court of Appeal in *R (Edward Bridges) v The Chief Constable of South Wales Police*[74] contains careful analysis of Data Protection Impact Assessments ('DPIA's') within public authority decision-making. The two grounds of appeal which are relevant for this paper are set out at paragraph 53, and can be summarised as follows:

- South Wales Police breached the Data Protection 2018 (which imports the GDPR) by failing to conduct an adequate DPIA;
- In particular, the DPIA was inadequate because it contained an error of law, in assuming that the right to privacy under Article 8 ECHR did not apply.
- South Wales Police also breached the PSED by conducting an inadequate Equality Impact Assessment, which did not sufficiently address the risks of indirect discrimination .[75]

These grounds were successful: the Court of Appeal held that the DPIA was insufficient, and that the risk of indirect discrimination was insufficiently recognised by South Wales Police. In short, the Court of Appeal concluded that insufficient efforts had been made by the police to investigate the potential for racial or gender bias in the software they had licenced,[76] leading to breaches of both data protection law and the public sector equality duty. This case thus illustrates the natural alignment of these obligations: had South Wales Police conducted a more thorough and legally accurate DPIA, they would have stood a better chance of identifying the risks of indirect discrimination on racial and gender grounds, and thus of complying with the Public Sector Equality Duty. This 'more thorough' DPIA could, I suggest, have been completed by aligning the DPIA and Equality Impact Assessment ('EIA') processes, to create what I term the 'DPIA+' model.

4.3 DPIA+ PSED = DPIA+

The advantages of combining the DPIA and PSED evaluation processes are potentially manifold, including the more ineffable benefit of drawing from two evolving schools of thought, which can both

[73] [2019] EWHC 2341 (Admin), [1].
[74] Note 19.
[75] Unlike a DPIA, which the GDPR makes mandatory in some circumstances, an Equality Impact Assessment is *not* a mandatory assessment under the Equality Act 2010, but it is a recommended means of ensuring compliance with the PSED. See note 37.
[76] Note 19, [201-202].

help refine broad concepts such as 'fairness.'[77] Prosaically, there is also the practical advantage of avoiding duplication of time and effort, where there is overlap. The two advantages considered in this subsection are 1) drawing together a single pool of stakeholders and 2) mutually reinforcing each other's requirement to investigate risk of discrimination.

4.3.1 Expanded Scope

The PSED directs attention to the risk of discrimination on the basis of 'protected characteristics,' which overlap with the GDPR's special categories of data without mirroring them entirely:

Equality Act 2010, s.149(7)	GDPR, Article 9(1)
Age	
Gender reassignment	
Sex	
Disability	[Data concerning health]
Pregnancy and maternity	[Data concerning health]
Race	Race or ethnic origin*
Religion and belief	Political opinions Religious or philosophical belief
Sexual orientation	Sex life or sexual orientation
	Trade union membership
	Genetic data
	Biometric identifiers

Table 1: Protected Characteristics under Equality and Data Protection Laws

The GDPR's category of 'data concerning health' is broader than any single protected characteristic, but the table (1) above suggests the areas of overlap. The GDPR is—perhaps—more aimed at novel forms of informatics, such as genetic and biometric data, than the demographic data that the Equality Act 2010 identifies, nonetheless, as potential loci of discrimination. The combined effect of the two amounts to a broader span of characteristics, which may still be helpful to consider together—for example, under a combined consultation process, to help identify groups at greater risk of harm. To use *Bridges* as an example, the identified risk of discrimination was on the basis of race and gender, and the latter could have been excluded from a DPIA for falling outside the scope of the GDPR's 'special categories' of personal data.

4.3.2 Duty to Investigate

In addition to the statutory language of s.149 EA 2010, the requirements of the PSED have been developed in judicial commentary now approved by the Supreme Court[78] in *R(Bracking) v Secretary of State for Work and Pensions*.[79] The courts have elaborated that the duty is proactive, substantive and may require active investigation:

The PSED must be fulfilled before and at the time when a particular policy is being considered. […]*The duty must be exercised in substance, with rigour, and with an open mind. It is not a question of ticking boxes. If the relevant material is not available, there will be a duty to acquire it and this will frequently mean that some further consultation with appropriate groups is required.*[80]

[77] In a different context, Gefenas et al argue that data protection should be considered in conjunction with research ethics on issues such as 'public interest,' as these separate aspects of compliance can become institutionally siloed. Eugenijus Gefenas, Jurate Lekstutiene, Vilma Lukaseviciene, et al, 'Controversies between regulations of research ethics and protection of personal data: informed consent at a cross-road' (2022) 25 Medicine Health Care and Philosophy 2, 23–30.
[78] *Hotak v Southwark LBC* [2015] UKSC 30, [2016] AC 811, at [73] (Lord Neuberger PSC).
[79] [2013] EWCA Civ 1345
[80] Note 19 [175].

The DPIA also has, to some extent, a consultation requirement in the form of Article 35(9):

Where appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of processing operations.

The crucial words, here, are '*where appropriate*.' There are many circumstances where the data controller has discretion to decide it is not 'appropriate' to consult with data subjects. In the case of public authorities, it is all too easy for public interest or security to be cited as reasons militating against transparency. In the context of the PSED, however, public authorities seem to have less leeway. In the words of the Court of Appeal judges in their combined *Bridges* judgment:

If the relevant material is not available, there will be a duty to acquire it and this will ***frequently*** *mean that some further consultation with appropriate groups is required* (emphasis added).[81]

Where a DPIA is conducted alongside consideration of the PSED, therefore, the GDPR's '*where appropriate*' caveat should be treated with more caution. These investigatory efforts do not end with what the ICO term 'participatory design' in their guidance on the use of AI,[82] and can also require expert assessment. In the case of *Bridges*, the Court of Appeal carefully reviewed the technical evidence on the risk of bias in AFR. While they did not make any finding that the specific software procured by South Wales Police posed a risk of bias, they noted the general risk of disproportionate error affecting Black people and women in AFR software. As such, the lack of investigation, and evidence that their software did *not* contain discriminatory bias, was sufficient to constitute a failure of the PSED.[83]

South Wales police would have had a reasonable degree of discretion as to the nature of this investigation, but one possibility would be a proxy assessment. A proxy analysis is one of the tools the ICO recommends as part of its AI guidance[84]—a test to see whether an algorithm falsely correlates a data feature (such as appearance on a watch list) with a protected characteristic (such as race or gender). Based on the Court's findings in *Bridges*, proxy assessment prior to the deployment (or even procurement) of new software by a public authority could have been a useful tool to investigate risk of discrimination under the DPIA and the PSED. Combining the two is, at least, an efficiency, but also an opportunity to draw from the guidance available under both statutes, and gain a fuller account of discrimination risk. As the Court of Appeal noted in their combined judgment:

[181] *We acknowledge that what is required by the PSED is dependent on the context and does not require the impossible. It requires the taking of reasonable steps to make enquiries about what may not yet be known to a public authority about the potential impact of a proposed decision or policy on people with the relevant characteristics, in particular for present purposes race and sex.*

The combined effect of DPIA and the PSED thus serves to mutually reinforce their requirements to proactively investigate the risk of bias, prior to an action being taken. Had South Wales Police considered the data protection and equality act duties holistically, these should have prompted investigation of the risks of indirect discrimination, which could in turn have prevented the omissions which the Court of Appeal ultimately found to be breaches of the PSED.

In short, the combination of the DPIA and the PSED is promising both in its resulting expansion of legal scope (the greater range of personal characteristics to be considered) and also in the mutual

[81] Ibid.
[82] Information Commissioner's Office, 'Guidance on AI and data protection', 'Annex A: Fairness' available from: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/annex-a-fairness-in-the-ai-lifecycle/ (accessed 17 March 2025).
[83] Note 19 [164-202].
[84] Note 82.

reinforcement of the respective investigative requirements. This combination would have been useful as a proactive consideration in the South Wales Police case discussed above, but this is not the only example of its utility.

To explore the potential benefit of combining equality and data protection law considerations, the next section returns to the main case study of this paper: the NHS Covid-19 data store.

5. The Covid-19 Data Store

5.1 The Beginning

The UK government announced the construction of the NHS Covid-19 data store in March 2020. A large part of the store was built, and put into operation, before an accompanying Data Protection Impact Assessment was published. [85] The Secretary of State's instructions to share data were first issued in March 2020, with the accompanying DPIA published in June 2020.[86] We do not know when the DPIA process for the data store was initiated, or what form it took. Nevertheless, healthcare providers were legally obliged to disclose patient information for a broad array of 'Covid-19 purposes,' before it was publicly established that the intended Covid-19 data store was compliant with data protection law.

As an initial point, which is crucial for this paper—as much impact assessment as possible should be conducted, and published— prospectively. Article 35(1) GDPR is clear that a DPIA should take place '*prior to the processing.*' Even in an urgent situation, therefore, a DPIA should be conducted and published at least in an initial form before the processing is legally mandated by central government. This means that health and social care providers, as well as the patient population, are given some assurance that the legally mandated disclosure of information has been reviewed from a data protection perspective, and at least a provisional analysis of harm has been conducted. This should not require any delay, but instead provide a complementary evaluative process, to ensure the limitations of the data can be understood and mitigated as the resource is built. In this sense, lawyers and data scientists have a joint interest in data quality, with early evaluation meeting the concerns of both.

5.2 The 'COPI' Notice

The legal origin of the data Store was an instruction from the Secretary of State for Health and Social Care, telling healthcare providers to share patient information for Covid-19 purposes. This instruction took the form of a 'COPI notice.' 'COPI' notices are made under the Control of Patient Regulations 2002. These notices can appear to provide a definitive "legal basis" for subsequent data processing—i.e., complete assurance that all processing is lawful. For example, the letter to NHS Digital states:

'*I consider this Notice is necessary so that NHS Digital can lawfully and efficiently disseminate confidential patient information*'[87]

This is true to the extent that a COPI notice provides a starting point for lawfulness. It ensures that the use of patient data is lawful, at least according to some of the applicable law. It gives healthcare providers a basis in law for the disclosure of patient data, meaning it should not breach their duty of confidence. It should also satisfy GDPR Article 6(1)(e), which requires public authorities to have a foundation in law when they process personal data. It does not, however, ensure that the processing

[85] Institute for Government, 'NHS Covid-19 Data Store and NHS National Data Platform Summary of a private roundtable' (2022) available from: https://www.instituteforgovernment.org.uk/sites/default/files/2023-02/nhs-covid-datastore.pdf (accessed 17 March 2025).
[86] The DPIA was first published 5 June 2020, see NHS England 'Data Protection Impact Assessment: NHS COVID-19 Data Store' available from: https://www.england.nhs.uk/publication/data-protection-impact-assessment-nhs-covid-19-data-store/ (accessed 17 March 2025).
[87] Letter from Secretary of State for Health and Social Care to Sarah Wilkinson, Chief Executive of NHS Digital, (17 March 2020) available from: https://digital.nhs.uk/coronavirus/coronavirus-covid-19-response-information-governance-hub/control-of-patient-information-copi-notice (accessed 17 March 2025).

will comply with other areas of law, such as the Public Sector Equality Duty under s.149 Equality Act. It is also no substitute for a Data Protection Impact Assessment ('DPIA'). If a subsequent DPIA identified a high risk of harm to data subjects, which could not be sufficiently mitigated, the processing would still breach the GDPR, and thus be unlawful.[88]

5.2 The Data Store DPIA

The responsibility for conducting a DPIA for the Data Store fell to the assumed sole data controller for the NHS Data Store: NHS England.[89] There is an inherent tension in the application of the GDPR, here, as the Regulation envisages that the controller will have a significant amount of say in the means and purposes of processing—either alone or '*jointly with others.*'[90] This assumption is less straightforward when the main architecture of the processing has already been determined by the Secretary of State through a COPI notice. This question should have been considered within the DPIA itself, as a foundational point of accountability (e.g., whether the government department which mandated the data processing is, in fact, a joint controller).

Otherwise, the DPIA for the Covid-19 data store published by NHS England —now updated to version 5.a[91]— provides a thorough, factual picture of the processing operations involved in the construction and operation of the Data Store. In doing so, it apparently succeeds in providing a systematic review of the intended processing; the first requirement of a DPIA under Article 35(7)(a) GDPR. Some consideration is also given to the necessity of the processing, per Article 35 (7) (b), with the document citing the need to centralise information for the government, and the superior quality of national records compared to regional ones.[92]

The element of the DPIA which is the focus of this paper— investigation of risks to data subjects—is less apparent. There are, essentially, four requirements of a DPIA under the GDPR:

> *a) a systematic description of the envisaged processing operations and the purposes of the processing*[…][93]
>
> *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 referred to in paragraph 1; 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.*

Satisfaction of a DPIA's second two requirements c) and d) is not evidenced by NHS England's published DPIA. Under 'assessment of risk to data subjects,' the document simply states *'(t)he risk assessment contains security information which will not be published*.'[94] This statement may not tell

[88] And, under Article 36 GDPR, this would have to be resolved with the Supervisory Authority (in this case, the Information Commissioner's Office).
[89] The DPIA (note 86, page 7) states that NHS England is the sole controller '*for any data legally shared with them under the NHS Control of Patient Information (COPI) Notice issued by the Secretary of State*' but this statement is peculiar given that the Secretary of State had already made an initial determination of the purposes of processing within the COPI notice.
[90] Per the definition in Article 4(7).
[91] Note 86.
[92] Ibid, pages 1 and 16.
[93] The GDPR's reference to 'legitimate interests' has been omitted here, as under Article 6(1)(f) GDPR public authorities cannot rely on this ground, and it is therefore out of scope for this paper.
[94] Note 86, page 18.

the full story, as NHS England has published a separate risk assessment within an Excel spreadsheet.[95] But, as a starting point it is concerning. By excluding the detail of the risk assessment, it omits the most important evaluative exercise of a DPIA, and does not feed an assessment of risk into the overall narrative of its deliberation. Assessment of risks to the rights and freedoms of individuals is a compulsory part of a DPIA under the GDPR,[96] and without it any account of '*measures envisaged to address the risks*'[97] does not make sense. By failing to meet two out of the four statutory requirements for a DPIA, the published version falls significantly short of evidencing the evaluative exercise required by the GDPR.

The risk assessment Excel spreadsheet, published separately, improves on the blank redaction within the DPIA.[98] Its content is appropriate, and identifies information security risks which will inevitably accompany the centralisation of NHS data. The risks to individuals listed in the spreadsheet are strongly inclined towards information security, however, and are not linked in any detail to the particular purpose of the Covid-19 data store, for example:

The context in which information is used or disclosed can change over time, leading to it being used for different purposes without people's knowledge.

This is probably true, but a little generic. The words 'Covid-19' do not feature at all in the risks identified in this spreadsheet, nor any reference to the risk of skewed analytics impacting service provision, nor whether this is a heightened risk in respect of demographics vulnerable to under-representation within NHS datasets. These risks are explained further in the following subsection, with the benefit of subsequent studies. Although these studies were conducted retrospectively, that does not mean the risks of discrepancy they investigated were impossible to anticipate before the Data Store was built. The Equality Act 2010, and the General Data Protection Regulation, set out a range of characteristics which are (respectively) protected, or special categories of information. As the GDPR itself states, they are special characteristics because they are recognised as potential loci of (indirect[99]) discrimination, or other threats to fundamental rights and freedoms.[100]

These risks were possible to anticipate in March 2020, and could not have been iteratively reviewed as the Store was constructed and operated. A (non-exhaustive) list of considerations this could have included is detailed below.

6. Risks of Bias in the Covid-19 Data Store

While there are no certainties as to whether the following factors made a difference in the Data Store's analytics, I suggest they would have merited consideration under a DPIA—particularly one reinforced by the Equality Act 2010:

6.1 Opt-outs

A constant source of potential bias in NHS data comes from the implementation of patients' right to opt out from secondary uses of health data. This can be a cause of concern for any policymakers using patient data to inform their decision-making. It is possible that some patient groups may opt out of secondary uses of their data at a higher rate, meaning they are under-represented in the final dataset

[95] NHS England, 'DPIA Risk Register' available from: dpia-nhs-covid19-risk-assessment-for-data-store.xlsx (live.com) (accessed 17 March 2025).
[96] Article 35.7 (c) GDPR.
[97] Article 35.7 (d) GDPR.
[98] Note 95.
[99] I.e. when an apparently neutral intervention disproportionately affects people with certain protected characteristics—see Hugh Collins and Tarunabh Khaitan 'Indirect Discrimination Law: Controversies and Critical Questions' in Hugh Collins and Tarunabh Khaitan (eds) *Foundations of Indirect Discrimination Law* (2018) (Oxford, Hart Publishing).
[100] Article 51 GDPR.

available for analysis. For example, an Equality Impact Assessment for a regional NHS Secure Data Environment programme devotes a significant proportion of its analysis to the impact of the national data opt-out.[101]

Any data shared under a COPI notice from the Secretary of State should have been subject to the national opt-out.[102] From the subsequent round table report, it seems the question of '*how to adhere to people opting out of their data being used by the health service'*[103] was one of the initial delaying factors in the construction of the store.

As well as implementing the opt-out, it would have been important to consider its potential impact on the datasets. This is not necessarily a straightforward exercise. As Tazare and colleagues point out, one of the key limitations created by the opt-out is that it prevents the use of data for research— even research on the impact of the opt-out itself. This means there is limited demographic information about the individuals who are missing from NHS research data.[104] Whether this data under-represents, or over-represents, certain demographics is thus difficult to establish. For example, the authors suggest that in 2013, opt-outs were higher among Black people than among White and Asian groups, but they were unable to complete a more up-to date-analysis as ethnicity data were not available.[105] They do also note, in relation to one database:

Studies based on Clinical Practice Research (CPRD) databases, which consist of de-identified data from a large proportion of UK practices and do not include people who have opted out, suggest that the database was broadly representative of the UK population by age, sex, area-based deprivation, and ethnicity in May 2021 and before. This suggests that the low proportions of opt-outs at this time, and small differences between demographic groups, had minimal impact on overall representativeness at this stage.

These findings are encouraging, but they seem only to relate to one patient database (the CPRD). The DPIA for the Covid-19 data store, on the other hand, lists 35 sources of information,[106] which may vary in their representativeness. It bears repetition that the stated aim for the data store was to create '*a single source of truth*.'[107] Those making decisions based on analytics within the store should have been warned about any gaps in their single source of truth. This is the kind of mitigating safeguard against indirect discrimination a DPIA could appropriately discuss: whether decision-makers should be given these caveats before allocating resources/ lifting restrictions based on these analytics.

6.2 Age

A major demographic at risk of indirect discrimination was made up of patients over the age of 80. Age was by far the greatest factor in the disparity of outcome of Covid-19, with patients over 80 being seventy times more likely to die following a positive test result than those under 40.[108] Clearly,

[101] Dominic Rowney and Suprasad Gavhane, 'North East and North Cumbria Secure Data Environment programme— Health data sharing: Equality Impact Assessment' (February 2024) available from: https://northeastnorthcumbria.nhs.uk/media/adbevmrg/ppie-equality-impact-assessment.pdf (accessed 17 March 2025).
[102] NHS England, 'Understanding the national data opt-out' (16 May 2023) available from: https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out (accessed 17 March 2025).
[103] Note 85, page 12.
[104] John Tazare, Alasdair D Henderson, Jessica Morley et al, 'NHS national data opt-outs: trends and potential consequences for health data research' (2024) 8 BJGP Open 3, 2024.0020.
[105] Ibid.
[106] Note 86, pages 4-5.
[107] Note 13.
[108] Note 14, page 10.

decisions about the response to Covid-19 would have a particularly significant impact on older patients as compared to people more likely to survive the virus.

In making decisions such as the placing or lifting of lockdown restrictions, or whether to focus clinical resources on care homes, the disproportionate impact of the decision on older people would have required consideration under the PSED. To do so properly, the risk needed to be represented as accurately as possible from the information available. If there was any doubt as to the representativeness of the data in the data store's dashboards about patients aged 80+ (for example), these should be flagged.

Public Health England's report (published in August 2020)[109] suggests that deaths from Covid-19 in care homes were under-reported in March-May 2020. This is the kind of informational gap which could have under-represented the scale of the problem for this patient group within the data store dashboards, and where careful margins of error would be a useful safeguard, before decisions are taken. A simplified illustration would be:

Risk	Harm	Mitigation
It is not certain that we have full, up-to date reports of the number of deaths from Covid-19 within care homes.	Care home residents are disproportionately older and disabled, and deaths within these groups may therefore be inadvertently minimised when resource allocation is considered.	Data limitations will be investigated and reviewed on an ongoing basis, and subject to external peer review. Data store dashboards to flag the current margin of error, which will be reviewed and updated as often as possible.

This is not to say that the ultimate policy outcomes would, or should, have been different. Decisions with disproportionate impact on older patients could still have been taken, and justified under the Equality Act 2010.[110] But the role of a DPIA is to prevent avoidable harm from data processing—such as harms stemming from unfair or inaccurate use of data. As such, the risk of older patients being under-represented within a decision-making resource should have been considered within an impact assessment.

6.3 Race

Race is another protected characteristic under the Equality Act 2010 which raises questions of representation within NHS data. Access to NHS services—and thus representation within NHS data—is a complex issue, driven by a number of factors. In the context of virtual patient cohorts, for example, the NHS itself suggests that '*insufficient ethnicity data capture, language and cultural barriers, and a misalignment between referral demographics and geographic profiles'* could be a reason for BAME under-representation.[111]

Likewise, if people from minority ethnic groups were less likely to access NHS care during the Covid-19 pandemic,[112] this too could have distorted the government's '*single source of truth'* about

[109] Note 14, page 57.
[110] Age being one of the protected characteristics for the purposes of the Equality Act 2010.
[111] Tammy Lovell, 'Black and ethnic minority people under-represented in virtual wards' (3 June 2024) available from: https://www.digitalhealth.net/2024/06/black-and-ethnic-minority-people-underrepresented-in-virtual-wards/ (accessed 17 March 2025).
[112] For example, due to immigration status, or linguistic/cultural barriers—See Public Health England report, note 14, page 40.

the prevalence of the disease in England. This is not a straightforward risk to assess—if anything, one London study suggested that minority ethnic patients were over-represented in hospital data, due to higher rates of hospitalisation.[113] But in 2020, Public Health England reported higher rates of deaths for people in England who were born outside Europe,[114] also citing potential barriers in accessing services.[115] If these patients were less likely to access services, they may in turn have been less likely to feature in the Big healthcare Data informing the datastore. Racial minorities could therefore constitute another patient group (or set of groups) whose representation within the dashboards merited scrutiny, so that their needs were not inadvertently downplayed.

6.4 Disability

The risk of under-representation of care home residents has already been discussed. Another group of disabled patients who could also have been considered, however, are those with a learning disability.

This group was not among those highlighted by Public Health England in their August 2020 report on outcomes, but subsequent studies showed an association between learning disability and a higher risk of death from Covid-19.[116] This raises questions of how well patients with a learning disability would be flagged within the government's analytics. Williamson and colleagues point out that only an estimated 23% of people with learning disabilities are included on the national learning disability register, meaning that the association between the disability and mortality could well be higher than reported.[117] This, they argue, is concerning for the evidence base informing decisions around vaccine prioritisation and other preventative measures for this group.

As above, this underlines the importance of iterative impact assessment that keenly targets gaps in any personal data used to inform government decision-making. It may not be realistic to expect NHS England to identify every patient group at higher risk of Covid-19 at the outset of a pandemic, but the public sector equality duty requires reasonable investigation of the risks of discrimination[118]—an investigation which should be iterative, and reviewed as fresh information comes to light.

6.5 The (Retrospective) DPIA+ for Covid-19

To reiterate: the above is not a criticism of any decisions taken in reliance on the Covid-19 data store, or a suggestion that any patient group suffered unlawful or unjustified impact from these decisions. The importance of the data store resource to inform decision-making is clear, and it is evident that a huge amount of work went into its urgent construction.[119] However, combining DPIA with the public-sector equality duty in relation to health data governance— i.e. the DPIA+ model for which this paper advocates— would have significantly reduced the risks of biases. This is not a call for an overhaul of existing processes, but rather for careful refinement. Evaluation of data bias, and discrimination risks within a DPIA process could be a complementary, supportive process that helps improve the quality of the information resource as it is designed and built. There is emerging recognition of how data accuracy as a data protection principle can help ensure fairness in algorithmic processing.[120] The

[113] Annastazia E. Learoyd, Jennifer Nicholas, Nicholas Hart, et al, 'Revisiting ethnic discrepancies in COVID-19 hospitalized cohorts: a correction for collider bias' (2023) 161 Journal of Clinical Epidemiology, 2023, Pages 94-103, https://doi.org/10.1016/j.jclinepi.2023.06.014.
[114] Ibid page 12, page 17.
[115] Ibid, page 40.
[116] Note 15.
[117] Ibid.
[118] Per the *Bracking* judgment, see note 79.
[119] Note 85.
[120] Elisabetta Biasin, 'Why accuracy needs further exploration in data protection' Proceedings of the 1st International Conference on AI for People: Towards Sustainable AI, CAIP 2021 (20-24 November 2021) Bologna, Italy.

evaluative model proposed here forms part of this alignment between technical accuracy and legal fairness.

7. Conclusion

This paper has advocated for a 'DPIA+' model of risk evaluation. This means that— prior to a new, large-scale programme of data processing— a public authority should evaluate the risks of harm to data subjects, both individually and at the level of any groups particularly at risk of discrimination. In the COVID-19 case study explored in this paper, the relevant juncture would be the Control of Patient Information, or 'COPI', notice handed down in 2020. This notice required healthcare organisations to share patient information before it was established that the ultimate purpose was compliant with data protection law, and would not lead to dangerously skewed analytics. In future, even urgent programmes of data processing could have a DPIA+ assessment to review the risks of processing— immediate and downstream— running concurrently with collection under a COPI notice. To do so should not mean delay, and would only serve to strengthen the programme of data processing by creating space to identify the biases and limitations of the available information, and considering how to mitigate the risk of skewed decision-making.

The case of *Bridges* illustrates that the DPIA+ model advocated for here has implications across the public sector, beyond healthcare. A combined consideration of data processing risk, and the public sector equality duty, could have saved South Wales Police from the dual breaches that were found to have taken place when they rolled out AFR software.

The risks of Big Data processing in the public sector are not set to abate. Even if a failure comparable to the Dutch welfare benefit scandal can be avoided, controversies about how these data resources are built— and legitimated in law— continue. The Covid-19 data store discussed in this paper has now been migrated into a longer-term NHS Federated Data Platform. From the reported commentary, some confusion appears to persist as to whether this platform requires a fresh COPI notice to ensure its lawfulness, or can be built using general statutory powers.[121] A DPIA has been published for this platform, but this document does not identify any of the risks of demographic-specific data bias which have been discussed in this paper.[122] Clearly, large-scale processing of citizens' data is set to remain a feature of public sector decision-making, and the DPIA's currently conducted are not capturing the full extent of the attendant risks. This paper highlights this missed opportunity, and advocates for a DPIA+ model to review risks to individuals and groups, and bring in considerations of the public equality duty. This is not mere red tape, but instead a chance to improve data-driven decision making in the public sector by testing the quality of the information used as a basis to make what can be, for some, life or death decisions.

[121] Lindsay Clark, 'Key aspects of Palantir's Federated Data Platform lack legal basis, lawyers tell NHS England' (5 September 2024) available from: https://www.theregister.com/2024/09/05/fdp_lacks_legal_basis/ (accessed 17 March 2025).

[122] NHS England, 'Overarching data protection impact assessment (DPIA) for the Federated Data Platform (FDP)' (14 November 2024) available from: https://www.england.nhs.uk/long-read/overarching-data-protection-impact-assessment-dpia-for-the-federated-data-platform-fdp/ (accessed 17 March 2025).

Conclusion

In this thesis, I have taken a 'privacy-based' approach to the regulation of secondary uses of patients' data. The overarching research question was whether my approach can help regulate these secondary uses. To answer this question, I have taken Article 8 ECHR as my central benchmark for defining 'privacy.' I have reviewed **Case Studies** illustrating the potential harms and benefits of secondary uses of patients' data, with a focus on three core concepts within Article 8:

Summary of Thesis Conclusions
Part 1. Article 8(1): Identification. Here, I concluded that *a person is identified by information if it can interfere with their right to private life.*
Part 2. Article 8(1): Private Life. I argued that *patients' identifiable data are 'private' if their secondary use can interfere with private life.*
Part 3. Article 8(2): Justification. I found that *secondary uses of patients' data should be justified along according to principles of proportionality and non-discrimination.*

Table 2: Summary of thesis conclusions

As summarised in the box above, the analysis conducted across the three Parts of this thesis has produced a revised account of these core aspects of Article 8, at least as they apply to secondary uses of patients' data. This clarified understanding helps to make English information law more internally consistent, and could also help improve the protection offered to patients when their data are re-used for purposes beyond their healthcare. My approach to each concept has thus enabled me to make both a *doctrinal* contribution to legal clarity, as well as a more *policy-orientated* contribution. The next subsections will explain the contribution of each Part.

4.1. Part 1: Identification

The clarified conception of identification I developed in Part 1 was: *a person is 'identified' by information when it interferes with their right to privacy under Article 8 ECHR.* This clarified conception of identification contributes to existing legal doctrinal work. Identification is a key concept in English data protection, confidentiality and privacy laws as it determines the (in)applicability of these laws. However, without a clear, internal conception of 'identity,' the idea of 'identification' is not a meaningful demarcator.

In the 3.profiling chapter,[196] I argued that interference with Article 8 ECHR would be a more coherent and purposive way to establish when such laws should apply. Basing the concept of

[196] Publication 3 (**thesis page 57**).

personal data on interference with Article 8 ECHR represents an alignment between privacy and data protection law. This adds to existing literature on the relationship between the two,[197] which is also discussed in the 7.DPIA+ article (in Part 3).[198]

This addition represents a valuable contribution, because it provides a central linchpin for the concept of personal data. Rather than searching for the concept of 'identification' in a list of potential identifiers and types of information (which are divorced from context, and can never be exhaustive[199]), it is more intellectually coherent to see the right to data protection as aligning closely with the right to privacy, meaning that information is personal if it has the capacity to impact on private life. Otherwise, the reader is left looking at a list of disparate potential identifiers (name, home address, ethnicity, mobile phone IP address, occupation etc.), wondering what essential quality these pieces of information have in common.

This definition has the advantage of conceptual clarity, but also of practical flexibility. If the concept of 'identification' is too nebulously defined, the scope of data protection, privacy and confidentiality laws is rendered ambiguous, as all these laws protect identified or identifiable individuals. At the same time, if the scope of 'identification' is drawn too broadly, or rigidly, this can prevent the benefit to patients associated with secondary uses of their anonymised NHS data.[200]

The benefits of flexibility inherent in my proposed approach are illustrated by the case studies I have focused on, including cell therapies (in the 2.anonymity article[201]), in the **Tissue Donation Case Study**, which require 'donor anonymity.' I also considered population-level research in the **Scientific Research Case Study** (in the 1.pseudonymisation article[202]), which does not need to scrutinise or evaluate any particular individual to make its findings. As the 3.profiling chapter[203] clarified, the capacity to scrutinise or evaluate a person as a unique human individual, and not just a data point within a larger pattern, is a key example of what would constitute an interference with Article 8 ECHR.

The flexibility of this approach helps make the concept of 'identification' a useful tool to regulate patients' data. By rejecting a 'data-centric' approach,[204] and instead rooting

---

[197] M. Tzanou, 'Data protection as a fundamental right next to privacy? "Reconstructing" a not so new right' (2013) 3 International Data Privacy Law 2, 88–99, 89; E. S. Dove, 'The EU General Data Protection Regulation: implications for international scientific research in the digital era' (2019) 46 Journal of Law, Medicine & Ethics 4, 1013-1030; O. Lynskey, 'Deconstructing data protection: the "added-value" of a right to data protection in the EU legal order' (2014) 63 International and Comparative Law Quarterly 3, 569-597; J. Kokott and C. Sobotta, 'The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 4, 222–228.

[198] Publication 7 (**thesis page 77**), internal page number 9.

[199] M. Elliot, E. Mackey, K. O'Hara and C. Tudor, *The Anonymisation Decision-Making Framework* (Manchester: UK Anonymisation Network, 2016). As discussed in the 1.pseudonymisation article (**thesis page 55**) internal page number 3.

[200] See W. N. Price and I. G. Cohen: 'it is important that we note assume privacy maximalism across the board is the way to go. Privacy underprotection and overprotection each create cognizable harms to patients both today and tomorrow.' 'Privacy in the age of medical big data' (2019) 25 Nature Medicine, 37-43.

[201] Publication 2, **thesis page 56**.

[202] Publication 1, **thesis page 55.**

[203] Publication 3, **thesis page 57**.

[204] Also rejected by Elliot et al (note 199).

identification within interference with Article 8 ECHR, this leaves the door open to a context-specific assessment. Hence the argument in the 2.anonymity article[205] that the identity/anonymity binary can be an important regulatory tool, which supports careful information management. The clarified conception of 'identification' in this thesis can thus be seen as a policy benefit for scientific research and tissue donation, as well as a doctrinal contribution.

In sum, *doctrinally*, Part 1 provides a definition of 'identity' for the purposes of the concept of personal data, which a) solidifies what can otherwise be a nebulous concept, and b) creates a link between the rights to data protection and privacy, by making interference with private life a common trigger for both. The *practical* implication of this definition is that it helps draw a meaningful distinction between a pseudonym and a profile— meaning that patients' information can be used anonymously, within secondary uses, as long as there is insufficient detail for them to be uniquely profiled.

4.2. Part 2: Private Life

Part 2 of this thesis argued that *patients' identifiable data are 'private' if their secondary use can interfere with private life.* In the 4.reasonable expectations article, I argued these data should be considered private by default when used for secondary purposes. Following the 'privacy-based approach' of my thesis, I used the term 'private' to mean within the proper scope of private and family life per Article 8.

In terms of legal doctrine, my definition of 'private life' is the other dimension of the alignment I have drawn between privacy and data protection. If personal data is defined as all information which interferes with Article 8 ECHR (as argued in Part 1), then conversely all personal data should be an aspect of private life. This reduces the need for additional considerations of 'reasonableness' or triviality, or any additional gloss on the concept of personal data for them to be considered private. This challenges both *dicta* and *ratio* in of some English judgments, which have suggested (*obiter dicta*) that 'trivial' health information like a is not serious enough to be considered private,[206] and that (*ratio decidendi*) patients should be circumstantially 'reasonable' in expecting their information to be treated as private and confidential.[207]

As well as bringing English caselaw into better alignment with that of the ECtHR, I suggested that there were also strong policy reasons for my challenge to the centrality of 'reasonable expectations' in English common law, particularly in terms of the clarity and accessibility of any interferences in patients' private lives. This argument was made in the con text of the **Immigration Case Study** and the **DeepMind Case Study**, both of which demonstr ated the inherent potential for patients' identifiable data to interfere with their private and fami ly lives, particularly when used for purposes other than their healthcare.

[205] Publication 2, **thesis page 56**.
[206] See Lady Hale's comments on the privacy interest in the reporting of a broken leg, or a cold. *Campbell v MGN Ltd* [2004] UKHL 22; [2004] 2 AC 457, [157].
[207] [2015] EWCA Civ 1034, Introduction, note 88.

As secondary uses of patients' data have proliferated, and the complexity of the data lifecycle has increased, it is questionable whether personal data held by public authorities can be 'trivial.' The increasing multiplicity of purposes for which the information could be used challenges the idea that information is incapable of impacting a patient, at least to the degree of individual scrutiny which the 3.profiling chapter[208] set as the baseline for interference with Article 8. The caselaw of the European Court of Human Rights also bears out the idea that the mere storage of personal data by public authorities engages Article 8 ECHR.[209]

This alignment between 'personal' and 'private' has practical value for secondary uses of patients' data. For example, the UK government has historically been uncertain as to whether the safeguards for patient data should extend only to 'confidential patient information,'[210] or to all personal data relating to patients. In 2018, for example, the Department of Health consulted on the National Data Opt-Out for England, seeking views on whether the opt-out should apply to all of patients' personal data, or just their 'confidential patient information' (i.e., 'CPI', or their clinical information).[211] The ultimate conclusion was that the opt-out should only apply to the latter, confidential patient information. However, following the findings of my thesis, this is a question I could revisit in further work.

In sum, Part 2 offers a *doctrinal* contribution, in its attempts to re-align the UK and the ECtHR construction of the scope of Article 8, by reducing the importance of 'reasonable expectations of privacy.' But it also offers a *policy* contribution, in as much as I suggest that there should be an onus on a Defendant (such as the NHS) to bring evidence to disprove the application of Article 8 ECHR. By default, therefore, secondary uses of patients' identifiable data should be justified under Article 8(2). This contribution on the scope of patients' rights therefore has a key implication for the way interference with these rights should be evaluated and justified.

4.3. Part 3: Justification

Finally, Part 3 of this thesis focused on Article 8(2), and concluded that *secondary uses of patients' data should be justified along according to principles of proportionality and non-discrimination*. Following my 'privacy-based' approach, I explored an understanding of justification that embraced multiple dimensions of Article 8 ECHR. As well as the value Article 8 places on individual autonomy, I also considered the importance of proportionality under Article 8(2), and the antidiscrimination principle linked to Article 8 by Article 14 ECHR. I explored how these principles could help regulate secondary uses of patients' data, with particular reference to the **Covid-19 Case Study,** but also with further exploration of the **Scientific Research Case Study.**

[208] Publication 3, **thesis page 57**.
[209] Introduction, note 4 [67].
[210] As defined in s.251(10) NHS Act 2006.
[211] Department for Health and Social Care, National Data Opt-Out Roundtable, (8 January 2018). No documents from this Consultation meeting were made public, but I attended this meeting with Professor Ruth Gilbert to discuss the implications of expanding the national opt-out to cover all of patients' personal data, and not just their confidential patient information ('CPI').

My expanded understanding of justification represents a contribution in both doctrinal and policy terms. Doctrinally, autonomy[212] and informational self-determination[213] have been prominent values discussed within the ethical and legal literature surrounding Big Data. Secondary uses of health data fit within this larger phenomenon of Big Data analytics, and so a 'privacy-based' approach could be interpreted as focused predominantly on informational autonomy, and patients' rights to control their data.

However, by taking Article 8 ECHR as my linchpin, I have embraced all its dimensions and nuances in my approach. As such, I have also considered how to justify interference with Article 8, if its values are conceptualised as including proportionality and non-discrimination. This would require measures other than allowing patients the ability to control their information (an objective aimed more at preserving their autonomy). In the '5.academic governance article,'[214] I explored the ways that Article 8(2) ECHR can expand the existing requirements of proportionality in the GDPR. In publication 7, I set out a 'DPIA+' model to incorporate the anti-discrimination principles of Article 14 within data protection evaluative processes.

The 'DPIA+' model, in particular, indicates the policy implications of my findings. If Article 8 ECHR should be considered as engaged, by default, when patients' identifiable information is used for secondary purposes, it is worth considering the multiple facets of its requirements. This may mean expanding the range of regulatory tools used to evaluate and justify data use. For example, data subject rights are a useful tool for promoting patients' informational autonomy within data protection, and these rights could be expanded to help individuals tackle the scale and complexity of Big Data Analytics.[215] But considerations such as equality of treatment between groups of patients require systemic overview of data processing.[216] This is beyond the scope of what an individual patient can achieve via (e.g.) a subject access request. Rather, these systemic considerations require NHS organisations to expand the parameters of their current evaluative practice. Hence my advocacy of a 'DPIA+' in publication 7.

In sum, by re-interpreting English information law in the light of the privacy interests enshrined in Article 8, the governance of secondary uses of patient data will benefit from improved doctrinal clarity and coherence, as well as better policy evaluation of 'Big' uses of patients' data. This is a doctrinal contribution, inasmuch as it helps to counter claims that the GDPR is too individualistic, or too broad in its exemptions, but also a policy contribution in the sense that it supports the development of practical tools, such as the DPIA+.

---

[212] Thesis Introduction, VI. C, 'Beyond Autonomy.'
[213] N. Purtova, 'Default entitlements in personal data in the proposed Regulation: Informational self-determination off the table… and back on again?' (2014) 30 Computer Law & Security Review 1, 6-24; J. C. Buitelaar, 'Child's best interest and informational self-determination: what the GDPR can learn from children's rights' (2018) 8 International Data Privacy Law 4, 293-308; J. C. Buitelaar, 'Post-mortem privacy and informational self-determination' (2017) 19 Ethics and Information Technology, 129-142.
[214] Publication 5, **thesis page 75.**
[215] S. Wachter and B. Mittelstadt, Introduction, note 76.
[216] E. Vayena and A. Blasimme, Introduction, note 79.

4.4. Overall Conclusion

The research question of this thesis was whether my 'privacy-based' approach, rooted in Article 8 ECHR, could help regulate secondary uses of patients' data. The overall conclusion is that this centring of Article 8, as a constitutional underpinning of the different strands of English information law, does indeed help identify commonalities between data protection, MOPI and confidentiality. The utility of this alignment lies partly in the doctrinal coherence it offers. But there is also a normative dimension: a re-affirmation that any patient whose information is held by the NHS should be entitled to privacy in the fullest sense that the ECHR envisages, comprising autonomy, dignity and anti-discrimination.

By re-grounding data protection, confidentiality and MOPI law in Article 8 ECHR, I have been able to form revised definitions which draw better lines of coherence between them: in particular, between the concepts of 'identification' (which is at the core of 'personal data,' in data protection law) and 'private' or 'confidential' information.' I have suggested the scope of these concepts should be drawn with ultimate reference to interference with private life under Article 8. This expands their scope, albeit in a flexible, impact-based way. It also changes how uses of these data should be justified. This means that, by default, consideration should be given to the potential for harm[217] to patients (or groups of patients).

The concepts I have examined in this thesis are not only used to regulate patients' data. A clarified understanding of identification, private life and justification could be useful across English information law, and my doctoral research could therefore be developed, in future work, to address the law at this broader level. Within this thesis, however, my contribution has clarified the scope of data protection, MOPI and confidentiality laws as they apply to secondary uses of patients' data, by using Article 8 ECHR to bridge the gap between them. In short: NHS patients' personal data should be considered private by default, and their secondary uses justified according to principles of proportionality, dignity and non-discrimination.

Article 8 ECHR has been an interest and concern in all my doctoral publications on secondary uses of patients' data. In the eight-year period in which these publications were written, I have consistently viewed it as a bedrock of English information law, which warrants exploration and emphasis in the context of secondary uses of data. My overarching contribution has thus evolved organically from this consistent point of emphasis. It is a foundational legal commitment within English law, which I have argued can bring coherence to the doctrine it influences. It is also my hope that it has helped bring coherence to this thesis, and to the legal doctrinal contribution it offers.

[217] Harm specifically in the sense of the violations of privacy, dignity and autonomy envisaged by Article 8 ECHR.

Bibliography

BOOKS

Aplin, T., L. Bently, P. Johnson et al, *Gurry on Breach of Confidence: The Protection of Confidential Information* (Oxford: Oxford University Press, 2012)

Beitz, C., *The Idea of Human Rights* (Oxford: Oxford University Press, 2009)

Chico, V. *Genomic Negligence: An Interest in Autonomy as the Basis for Novel Negligence Claims Generated by Genetic Technology* (Abingdon: Routledge-Cavendish, 2011)

Dennis, I., *The Law of Evidence* (London: Sweet & Maxwell, 2024)

Elliot, M., E. Mackey, K. O'Hara and C. Tudor, *The Anonymisation Decision-Making Framework* (Manchester: UK Anonymisation Network, 2016)

Emson, R., *Evidence* (Basingstoke: Palgrave Macmillan, 1999)

Hickman, T., *Public Law after the Human Rights Act* (Oxford: Hart Publishing, 2010)

Jay, R. *Guide to the General Data Protection Regulation: A Companion to Data Protection Law and Practice* (London: Sweet & Maxwell, 2017)

Kaye, J., S. Gibbons, C. Heeney A. Smart and M. Parker,, *Governing Biobanks – Understanding the Interplay between Law and Practice* (Oxford: Hart, 2012)

Keikhosroki, P (ed), *Big Data Analytics for Healthcare : Datasets, Techniques, Life Cycles, Management, and Applications* (London: Academic Press, 2022)

Laurie, G. *Genetic Privacy: A Challenge to Medico-Legal Norms* (Cambridge: Cambridge University Press, 2002)

Mantelero, A. *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI.* (The Hague: Springer, 2022)

O'Hara, K., *The seven veils of privacy: how our debates about privacy conceal its nature* (Manchester*:* Manchester University Press, 2023)

Sullivan, C., *Digital Identity: An Emergent Legal Concept* (Adelaide: University of Adelaide Press, 2011)

Taylor, L., L. Floridi and B. van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Cham: Springer, 2017)

Taylor, M., *Genetic Data and the Law: A Critical Perspective on Privacy Protection* (Cambridge: Cambridge University Press, 2012)

Toulson, R. and C. Phipps, *Confidentiality* (London: Sweet & Maxwell, 2012)

Walton, D., *Burden of Proof, Presumption and Argumentation* (Cambridge: Cambridge University Press, 2014)

Young, A., *Democratic Dialogue and the Constitution* (Oxford: Oxford University Press, 2017)

CHAPTERS IN A BOOK

Dove, E., ‘Misuse of private information and the common law right of privacy: a new frontier in biomedicine?’ in Dove (ed) *Confidentiality, Privacy and Data Protection in Biomedicine* (Abingdon: Routledge, 2024), 194-231

De Hert, P. and H. Lammerant, ‘Predictive Profiling and its Legal Limits: Effectiveness Gone Forever?’ in van der Sloot, Broeders and Schrijvers (eds) *Exploring the Boundaries of Big Data* (2016) (Amsterdam: Amsterdam University Press), 145-167

Elliot, M. and E. Mackey, ‘The Social Data Environment’ in O’Hara, Nguyen and Haynes (eds) *Digital Enlightenment Yearbook: Social Networks and Social Machines, Surveillance and Empowerment* (Amsterdam: IOS Press, 2014), 253-264

Evans, B. J., ‘Big Data and Individual Autonomy in a Crowd’ in Cohen, Fernandez, Vayena and Gasser (eds) *Big Data, Health Law, and Bioethics* (Cambridge: Cambridge University Press, 2018), 19-29

Dumortier, J. and P. Mahault, ‘European-Wide Big Health Data Analytics under the GDPR’ in Tzanou (ed) *Health Data Privacy under the GDPR: Big Data Challenges and Regulatory Responses* (Oxford: Routledge, 2021), 56–70

Forgó, N., S. Hänold and B. Schütze, ‘The Principle of Purpose Limitation and Big Data Perspectives in Law, Business and Innovation’ in Corrales, Fenwick and Forgó (eds) *New Technology, Big Data and the Law* (Singapore: Springer, 2017), 17-42

Mackor, A., ‘Explanatory Non-Normative Legal Doctrine, Taking the Distinction between Theoretical and Practical Reason Seriously’ in Van Hoeke (ed) *Methodologies of Legal Research: Which Kind of Method for What Kind of Discipline?* (London: Bloomsbury Publishing, 2013), 45-70

Minkkinen, P., ‘Critical legal “method” as attitude’ in Watkins (ed) *Research Methods in Law* (London: Routledge, 2017), 146-169

van Gestel, R., H. Micklitz and E. Rubin, ‘Should Doctrinal Legal Scholarship Be Abandoned?’ in van Gestel, Micklitz and Rubin (eds) *Rethinking Legal Scholarship: A Transatlantic Dialogue* (Cambridge: Cambridge University Press, 2017), 205–398

JOURNAL ARTICLES

Aidinlis, A., ‘The Right to be Forgotten as a Fundamental Right in the UK after Brexit’ (2020) 25 Communications Law 67

Aitken, M., J. de St. Jorre, C. Pagliari et al, ‘Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies’ (2016) 17 BMC Medical Ethics 73

Aldhouse, F., ‘Anonymisation of personal data: a missed opportunity for the European Commission’ (2014) 30 Computer Law & Security Review 4, 403-418

Allen, A., ‘Privacy-as-Data Control: Conceptual, Practical and Moral Limits of the Paradigm’ (2000) 32 Connecticut Law Review, 861-875

Andrew, J. and M. Baker, ‘The General Data Protection Regulation in the Age of Surveillance Capitalism’ (2019) 168 Journal of Business, 565–578

Arvind, T. and A. McMahon., ‘Responsiveness and the Role of Rights in Medical Law: Lessons from *Montgomery*’ (2020) 28 Medical Law Review 3, 445–477

Ballantyne, A., ‘How should we think about clinical data ownership?’ (2020) 46 Journal of Medical Ethics 5, 289-294

Banja, J., ‘Reasonable Persons, Autonomous Persons, and Lady Hale: Determining a Standard for Risk Disclosure’ (2020) 50 The Hastings Center Report 2, 25-34

Becker, R., A. Thorogood, J. Ordish, and M Beauvais, ‘COVID-19 Research: Navigating the European General Data Protection Regulation’ (2020) 22 Journal of Medical Internet Research 8, e19799

Bell, J., M. Mourby and J. Kaye, ‘Contractual Mechanisms for Securing the Public Interest in Data Sharing in Public-Private Health Research Partnerships’ (2023) 20 script*ed* 2, 325-351

Berberich, M. and M. Steiner, ‘Blockchain Technology and the GDPR—How to Reconcile Privacy and Distributed Ledgers?’ (2016) 2 European Data Protection Law Review 3, 422-426

Birnhack, M., ‘In Defence of Privacy-as-Control (Properly Understood) (2025) 65 Jurimetrics (forthcoming)

Black, J., ‘Regulatory Conversations’ (2002) 29 Journal of Law and Society 1, 163-96

Bolognini, L. and C. Bistolfi, ‘Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from Directive 95/46/EC to the new EU General Data Protection Regulation’ (2017) 33 Computer Law and Security Review 2, 171-181

Borgesius,F., ‘Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation’ (2016) 32 Computer Law and Security Review 2, 256-271

Buitelaar, J. C., ‘Post-mortem privacy and informational self-determination’ (2017) 19 Ethics and Information Technology, 129-142

Buitelaar, J. C., ‘Child’s best interest and informational self-determination: what the GDPR can learn from children’s rights’ (2018) 8 International Data Privacy Law 4, 293-308

Carter, P., G. Laurie and M. Dixon-Woods, ‘The social licence for research: why care.data ran into trouble’ (2015) 41 J Med Ethics 5, 404-9

Cohen, J., ‘Turning Privacy Inside Out’ (2019) 20 Theoretical Inquiries in Law 1, 1-22

Cohen, G., T. Coan, M. Ottey et al, ‘Sperm donor anonymity and compensation: an experiment with American sperm donors’ (2016) 3 J Law Biosciences 3, 468-488

Delacroix, S. and N. Lawrence, ‘Bottom-up Data Trusts: Disturbing the “One Size Fits All” Approach to Data Governance’ (2019) 9 International data privacy law 4, 236–252

Doerr, M. and S. Meeder, ‘Big Health Data Research and Group Harm: The Scope of IRB Review’ (2022) 44 Ethics and Human Research 4, 34-38

Dove, E., ‘The EU General Data Protection Regulation: implications for international scientific research in the digital era’ (2019) 46 Journal of Law, Medicine & Ethics 4, 1013-1030

Dove, E., ‘Confidentiality, public interest, and the human right to science: when can confidential information be used for the benefit of the wider community?’ (2023) 10 Journal of Law and the Biosciences 1, 1-53

Dove, E., S. Kelly, F. Lucivero et al, ‘Beyond Individualism: Is there a place for relational autonomy in clinical practice and research?’ (2017) 12 Clinical Ethics 3, 150–165

Ducato, R., ‘Data protection, scientific research, and the role of information’ (2020) 37 Computer Law & Security Review 105412

El Emam, K. and C. Álvarez, ‘A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques’ (2015) 5 International Data Privacy Law 1, 73-87

Elliot, M. and A. Dale, ‘Scenarios of attack: the data intruder’s perspective on statistical disclosure risk’ (1999) 14 Netherlands Official Statistics, 6-10

Elliot, M., K. O’Hara, C. Raab et al, ‘Functional anonymisation: Personal data and the data environment’ (2018) 34 Computer Law & Security Review 2, 204-221

Floridi, L., ‘Open Data, Data Protection, and Group Privacy’ (2014) 27 Philosophy & Technology 1, 1-3

Froomkin, M., ‘Big Data: Destroyer of Informed Consent’ (2019) 18 Yale Journal of Health, Policy, Law and Ethics 3, 27-54

Gillon, R., ‘Ethics Needs Principles—Four Can Encompass the Rest—and Respect for Autonomy Should Be “First among Equals”’ (2003) 29 Journal of Medical Ethics 5, 307-312

Hays, R. and G. Daker-White, ‘The care.data consensus? A qualitative analysis of opinions expressed on Twitter’ (2015) 15 BMC Public Health 1, 838

Heri, C., ‘Deference, Dignity and “Theoretical Crisis”: Justifying ECtHR Rights Between Prudence and Protection’ (2024) 24 Human Rights Law Review 1, 1-19

Iusmen, I. and J. Boswell, ‘The Dilemmas of Pursuing ‘Throughput Legitimacy’ through Participatory Mechanisms‘ (2016) 40 West European Politics 2, 459–78

Johnson, T., K. Kollnig and P. Dewitte, ‘Towards responsible, lawful and ethical data processing: patient data in the UK’ (2022) 11 Internet Policy Review 1, 1-25

Kaminski, M. and G. Malgieri, ‘Algorithmic impact assessments under the GDPR: producing multi-layered explanations’ (2021) 11 International Data Privacy Law, 2, 125–144

Kaye, J. S. Terry, E. Juengst et al, ‘Including all voices in international data-sharing governance’ (2018) 12 Human Genomics 13

Kish, L and E. Topol., ‘Unpatients-why patients should own their medical data’ (2015) 33 Nature Biotechnology 9, 921-4

Kokott, J. and C. Sobotta, ‘The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR’ (2013) 3 International Data Privacy Law 4, 222–228

Laurie, G. and L. Stevens, ‘Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom’ (2016) 43 Journal Law and Society 3, 360-392

Lynskey, O., ‘Deconstructing data protection: the “added-value” of a right to data protection in the EU legal order’ (2014) 63 International and Comparative Law Quarterly 3, 569-597

Mackey, E. and M. Elliot, ‘Understanding the data environment’ (2013) 20 XRDS: The ACM Magazine for Students 1, 36-39

Macmillan, K., ‘Baby Steps’ (2008) 13 Communications Law 3, 72-75

Mahieu, F., W. Decleer, K. Osmanagaoglu et al, ‘Anonymous sperm donors' attitude towards donation and the release of identifying information’ (2019) 36 Journal of Assisted Reproduction and Genetics, 2007-2016

Mantelero, A., ‘AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment’ (2018) 34 Computer Law & Security Review 4, 754–772

Mantelero, A. and M. Esposito, ‘An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems’ (2020) 41 Computer Law & Security Review 105561

McCartney, M., ‘Care.data: why are Scotland and Wales doing it differently?’ (2014) 20 British Medical Journal 348, 1702

McMahon, A., A. Buyx, and B. Prainsack, ‘Big Data Governance Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond’ (2020) 28 Medical law review 1, 155–182

Metcalfe, C., R. Martin, S. Noble, et al, ‘Low risk research using routinely collected identifiable health information without informed consent: encounters with the Patient Information Advisory Group’ (2008) 34 Journal of Medical Ethics 37-40

Minow, M., ‘Archetypal Legal Scholarship: A Field Guide’ (2013) 63 Journal of Legal Education 1, 65-69

Moreham, N., ‘Conversations with the common law: Exposure, privacy and societal change’ (2021) 52 Victoria University of Wellington Law Review 3, 563–577

Mourby, M., J. Doidge, K. Jones et al, ‘Health Data Linkage for UK Public Interest Research: Key Obstacles and Solutions’ (2019) 4 International journal of population data science 1

Nicholson Price, W. and I. G. Cohen, ‘Privacy in the age of medical big data’ (2019) 25 Nature medicine 1, 37-43

O’Brien, B., ‘Care.data: how Northern Ireland is doing it’ (2014) 3 British Medical Journal 348, 2380

Ohm, P., ‘Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization’ (2010) 57 UCLA Law Review, 1701-1777

Powles, J. and H. Hodson, ‘Google DeepMind and healthcare in an age of algorithms’ (2017) 7 Health Technology, 351-367

Presser, L. M. Hruskova, H. Rowbottom and J. Kancir, ‘Care.data and access to UK health records: patient privacy and public trust’ (2015) Technology Science, 2015081103

Purtova, N., ‘Default entitlements in personal data in the proposed Regulation: Informational self-determination off the table… and back on again?’ (2014) 30 Computer Law & Security Review 1, 6-24

Purtova, N., ‘The law of everything. Broad concept of personal data and future of EU data protection law’ (2018) 10 Law, Innovation and Technology 1, 40-81

Pormeister, K., ‘Genetic data and the research exemption: is the GDPR going too far?’ (2017) 7 International Data Privacy Law 2, 137-146

Price, W. N. and I. G. Cohen, ‘Privacy in the age of medical big data’ (2019) 25 Nature Medicine, 37-43

Richards, N. and D. Solove, ‘Privacy’s Other Path: Recovering the Law of Confidentiality’ (2007) 96 Georgetown Law Journal, 123-182

Ristevski, B. and M. Chen, ‘Big Data Analytics in Medicine and Healthcare’ (2018) 15 Journal Integrative Bioinformatics 3, 20170030

Rocher, L., J Hendrickx and Y. de Montjoye, ‘Estimating the success of re-identifications in incomplete datasets using generative models’ (2019) 10 Nature Communications 3069

Saunders, B., J. Kitzinger and C. Kitzinger, ‘Anonymising interview data: challenges and compromise in practice’ 15 Qualitative Research 5, 616-632

Schlegel, D. R. and G. Ficheur, ‘Secondary Use of Patient Data: Review of the Literature Published in 2016’ (2017) 26 Yearbook of Medical Informatics 1, 68-71

Schwartz, P., ‘Internet Privacy and the State’ (2000) 32 Connecticut Law Review, 815-859

Solove, D., ‘A taxonomy of privacy’ (2006) 154 University of Pennsylvania Law Review 3, 477–564

Solove, D., ‘Privacy self-management and the consent dilemma’ (2013) 126 Harvard Law Review 7, 1880–1903

Spadaro, A., ‘Covid-19: Testing the Limits of Human Rights’ 11 European Journal of Risk Regulation 2, 317-325

Spindler, G. and P. Schmechel, ‘Personal Data and Encryption in the European Data Protection Regulation’ (2016) 7 JIPTEC 2, 163-177

Stalla-Bourdillon, S. and A. Knight, ‘Anonymous Data v. Personal Data – False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data’ (2016) 34 Wisconsin International Law Journal 2, 284-322

Stevens, L., ‘The Proposed Data Protection Regulation and its Potential Impact on Social Sciences Research in the UK’ (2015) 1 European Data Protection Law Review 2, 97-112

Sterckx, S., V. Rakic, J. Cockbain and P. Borry, ‘“You hoped we would sleep walk into accepting the collection of our data": controversies surrounding the UK care.data scheme and their wider relevance for biomedical research’ (2016) 19 Medicine, Health Care and Philosophy 2, 177-190

Taghreed, J., ‘The UK’s National Programme for IT: Why was it dismantled?’ (2017) 30 Health Services Management Research 1, 2-9

Taylor M. and J. Wilson, ‘Reasonable Expectations of Privacy and Disclosure of Health Data’ (2019) 27 Medical Law Review 3, 432–460

Tzanou, M., ‘Data protection as a fundamental right next to privacy? “Reconstructing” a not so new right’ (2013) 3 International Data Privacy Law 2, 88–99

Vayena, E. and A. Blasimme, ‘Health Research with Big Data: Time for Systemic Oversight.’ (2018) 46 The Journal of law, medicine & ethics 1, 119-129

Wachter, S., ‘Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR’ (2018) 34 Computer Law and Security Review 3, 436-449

Watcher, S., ‘Data protection in the age of big data’ (2019) 2 Nature Electronics, 6-7

Wachter, S. and B. Mittelstadt, ‘A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI’ (2019) 2 Columbia Business Law Review, 494-620

ONLINE SOURCES

Ada Lovelace Institute, ‘Algorithmic impact assessment: user guide’ (8 February 2022) https://www.adalovelaceinstitute.org/resource/aia-user-guide/

Ada Lovelace Institute, ‘Exit through the App Store?’ (20 April 2020) https://www.adalovelaceinstitute.org/our-work/covid-19/covid-19-exit-through-the-app-store/

Academy of Medical Royal Colleges, ‘Disclosing personal demographic data: The public interest. Report of a seminar held at the Academy of Medical Royal Colleges’ (June 2021) https://www.aomrc.org.uk/wp-content/uploads/2021/06/Disclosing_personal_demographic_data_0621.pdf

Agencia Española Protección Data & European Data Protection Supervisor, ‘10 Misunderstandings Related to Anonymisation’ (2021) https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf

British Medical Association, ‘Brief for Annual representatives meeting, The ‘hostile environment’ (2021) https://www.bma.org.uk/media/4561/bma-arm-briefing-hostile-environment-arm2021.pdf

Burgess, M., ‘The law is nowhere near ready for the rise of AI-generated fake porn’ (27 January 2018) https://www.wired.co.uk/article/deepfake-app-ai-porn-fake-reddit

Clark, L., ‘Key aspects of Palantir's Federated Data Platform lack legal basis, lawyers tell NHS England’ (5 September 2024) https://www.theregister.com/2024/09/05/fdp_lacks_legal_basis/

Clarke, B., ‘Researchers: Anonymized data does little to protect user privacy’ (30 July 2019) https://thenextweb.com/news/anonymized-data-does-little-to-protect-privacy

CNIL, ‘Scientific research (excluding health): challenges and advantages of anonymization and pseudonymization’ (13 January 2022) https://www.cnil.fr/fr/recherche-scientifique-hors-sante/enjeux-avantages-anonymisation-pseudonymisation

Denham, E., Letter to Sir David Sloman (3 July 2017) https://ico.org.uk/media/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf

Department for Constitutional Affairs, Public Sector Data Sharing: Guidance on the Law (November 2003) https://www.mentalhealthlaw.co.uk/media/Data_sharing_legal_guidance.pdf

Department for Digital, Culture, Media and Sport, ‘Explanatory Notes to the Data Protection Bill 2017-19’ https://publications.parliament.uk/pa/bills/lbill/2017-2019/0066/18066en.pdf

Evans, H., ‘Using data in the NHS: the implications of the opt-out and GDPR’ (24 May 2018) https://www.kingsfund.org.uk/insight-and-analysis/long-reads/using-data-nhs-gdpr

European Commission Directorate-General for Health and Food Safety, ‘Question and Answers on the interplay between the Clinical Trials Regulation and the General Data Protection Regulation’ (2019) https://health.ec.europa.eu/document/download/c3042973-b36d-4094-a1fb-a6fc980f065e_en?filename=qa_clinicaltrials_gdpr_en.pdf

European Commission, Directorate-General for Health and Food Safety, ‘Assessment of the EU Member States’ rules on health data in the light of the GDPR’ (12 February 2021) https://health.ec.europa.eu/system/files/2021-02/ms_rules_health-data_en_0.pdf

European Data Protection Board, ‘EDPB letter to the European institutions on the privacy and data protection aspects of a possible digital euro’ (2021) https://edpb.europa.eu/system/files/2021-07/edpb_letter_out_2021_0112-digitaleuro-toep_en.pdf

European Medicines Agency, ‘Note for Guidance on Structure and Content of Clinical Study Reports’ (July 1996) https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-3-structure-content-clinical-study-reports-step-5_en.pdf

European Medicines Agency, ‘External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human

use' (15 October 2018) https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-3.pdf

Foxglove, 'Legal action launched: no legal basis for the £330 million Palantir NHS Federated Data Platform' (30 November 2023) https://www.foxglove.org.uk/2023/11/30/legal-action-palantir-nhs-federated-data-platform/

General Medical Council, 'Confidentiality: good practice in handling patient information' (January 2017) https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/confidentiality

Hodson, H., 'Revealed: Google AI has access to huge haul of NHS patient data' (29 April 2016) https://institutions.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/

Home Office, 'Home Office in the media: 10 May 2018' (10 May 2018) https://homeofficemedia.blog.gov.uk/2018/05/10/home-office-in-the-media-10-may-2018/

Home Office, Department of Health and NHS Digital, 'Memorandum of Understanding between Health and Social Care Information Centre and the Home Office and the Department of Health' (1 January 2017) https://assets.publishing.service.gov.uk/media/5c4f2103e5274a492e19de96/MoU_between_HSCIC__Home_Office_and_DH.pdf

Information Commissioner's Office, 'Anonymisation: managing data protection risk code of practice' https://ico.org.uk/media/1061/anonymisation-code.pdf

Linklaters LLP, 'Audit of the acute kidney injury detection system known as Streams' (17 May 2018, updated 7 June 2018) https://www.royalfree.nhs.uk/application/files/1516/9721/4007/Streams_Report.pdf

Matrix Chambers, 'Migrants' Rights Network granted permission for judicial review of patient data-sharing agreement between NHS Digital and the Home Office' (1 March 2018) https://www.matrixlaw.co.uk/news/migrants-rights-network-granted-permission-legally-challenge-data-sharing-agreement-nhs-digital-home-office/

McCrudden, C. and Sacha Prechal, 'The Concepts of Equality and Non-Discrimination in Europe: A practical approach' (2009) https://ec.europa.eu/social/BlobServlet?docId=4553

Medical Research Council, 'Identifiability, anonymisation and pseudonymisation: Guidance note 5' (September 2019) https://www.ukri.org/wp-content/uploads/2021/11/MRC-291121-GDPR-Identifiability-Anonymisation-Pseudonymisation.pdf

Melvin, G., 'We must listen to the public on GP data' (26 May 2022) https://digital.nhs.uk/blog/data-points-blog/2022/we-must-listen-to-the-public-on-gp-data

Miller, C., 'Would you recognise yourself from your data?' (29 May 2019) https://www.bbc.co.uk/news/technology-48434175

National Data Guardian for Health and Care, 'Review of Data Security, Consent and Opt-Outs' (June 2016)

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF

Nursing & Midwifery Council, ‘Professional standards of practice and behaviour for nurses, midwives and nursing associates’ (10 October 2018) https://www.nmc.org.uk/globalassets/sitedocuments/nmc-publications/nmc-code.pdf

NHS Digital, ‘A guide to confidentiality in health and social care: references. Section 4: Human Rights Act provisions’ (13 January 2022), https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/codes-of-practice-for-handling-information-in-health-and-care/a-guide-to-confidentiality-in-health-and-social-care/hscic-guide-to-confidentiality-references/section-4

NHS Digital, ‘Patient data and confidential patient information’ (12 October 2023): https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out/confidential-patient-information.

NHS England, ‘The NHS AI Lab’ https://transform.england.nhs.uk/ai-lab/

NHS England, ‘NHS artificial intelligence (AI) giving patients better care and support’ (12 December 2024) https://www.england.nhs.uk/2024/12/nhs-ai-giving-patients-better-care-and-support/#:~:text=The%20NHS%20is%20using%20AI,reducing%20demand%20on%20pressured%20A%26Es

NHS England, ‘Data Use Registers’ (2 July 2025) https://digital.nhs.uk/services/data-access-request-service-dars/data-uses-register#data-uses-registers

NHS England, ‘Federated Data Platform: information governance framework’ (7 August 2024)

https://www.england.nhs.uk/long-read/federated-data-platform-information-governance-framework/#:~:text=No%20Personal%20Data%20will%20be,259%20of%20the%20Health%20and

NHS England, ‘Invoice validation’ https://www.england.nhs.uk/ig/in-val/

NHS England, ‘NHS Digital and NHS England complete merger’ (1 February 2023) https://www.england.nhs.uk/2023/02/nhs-digital-and-nhs-england-complete-merger/

Niker, F., ‘An Interview with Baroness Onora O’Neill (Beyond the Ivory Tower series)’ (6 January 2020) http://justice-everywhere.org/governance/an-interview-with-baroness-onora-oneill-beyond-the-ivory-tower-series/

Ó Cathaoir, K., E. Gefenas, M. Hartlev, M. Mourby and V. Lukaseviciene, ‘EU-STANDS4PM report: Legal and ethical review of in silico modelling’ (March 2020) https://www.eu-stands4pm.eu/lw_resource/datapool/systemfiles/elements/files/AA77832F664661DBE0537E695E8689E3/current/document/WP3_March2020_D3-1_V1_public.pdf

Office for AI, ‘Algorithm: Showcasing Artificial Intelligence in Great Britain and Northern Ireland’ (September 2021)

https://assets.publishing.service.gov.uk/media/615301e2e90e077a38dc5c7b/ALGORITHM__Single_page_spread_.pdf

Public Health England, 'Data-sharing MoU between NHS Digital and Home Office: call for evidence' (26 March 2018) https://www.gov.uk/government/calls-for-evidence/data-sharing-mou-between-nhs-digital-and-home-office-call-for-evidence#FULL-PUBLICATION-UPDATE-HISTORY

Pyper, D., 'The Public Sector Equality Duty and Equality Impact Assessments' (2020) available from: https://researchbriefings.files.parliament.uk/documents/SN06591/SN06591.pdf

Ravindranath, M., 'How your health information is sold and turned into "risk scores"' (2 March 2019) https://www.politico.com/story/2019/02/03/health-risk-scores-opioid-abuse-1139978

Taranto, L. and P. Garcia, 'Medical Research Council Advises on How to Anonymise Information for Research Purposes' (16 October 2019) https://www.hldataprotection.com/2019/10/articles/international-eu-privacy/medical-research-council-advises-on-how-to-anonymise-information-for-research-purposes/

Thompson, B., 'Analysis: Research and the General Data Protection Regulation,' (July 2016) https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf

UK Statistics Authority, 'Research Code of Practice and Accreditation criteria' (July 2018) https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2018/08/COP_Research-and-Accreditation_A4.pdf

Understanding Patient Data, 'Identifiability Demystified' (5 April 2017) https://understandingpatientdata.org.uk/sites/default/files/2017-07/Identifiability%20briefing%205%20April.pdf

Wollaston, S., 'Letter from Dr Sarah Wollaston MP to Sarah Wilkinson' (29 January 2018): https://publications.parliament.uk/pa/cm201719/cmselect/cmhealth/Correspondence/Wilkinson-2018-01-29.pdf

World Health Organization, 'Ethics and governance of artificial intelligence for health' (28 June 2021) https://www.who.int/publications/i/item/9789240029200