A Socio-Legal Exploration of Approaches Towards Secondary Uses of Administrative Data in Social Science Research.



Hannah Smith

St Cross College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
MT 2021

99,500 words

ABSTRACT

Governments worldwide are increasingly seeking to capitalise on technological advances in the collecting, linking, and analysis of ever greater sets of data. By applying novel analytical techniques to previously collected administrative data, governments and researchers seek to shed light on the causes of societal challenges and provide a robust evidence base for policymaking.

However, these advances foster new opportunities and risks, particularly to citizens' privacy interests. Recent legislation has grappled with this modern data processing landscape by protecting individuals without unduly constraining the potential of such developments. Nevertheless, public backlashes to legal data-sharing initiatives and a wider societal unease towards the use of citizens' data question the ability of the law to secure societal acceptance in this context. Future data-sharing initiatives and the wider legitimacy of the law are put at risk if they diverge from citizens' views as to what is acceptable regarding data use. But how do such divergences arise, and how can we respond?

An analysis of the General Data Protection Regulation and the Digital Economy Act 2017 provides insights into the influences that shaped these legislative outcomes. Understanding the law as the product of a fiercely contested process highlights the underlying motivations that shaped the law and the effects that arise as a result. A mixed methods study provides rich data on citizens' views towards the appropriate secondary uses of administrative data in research. This enables a comparison between the legislative approaches and citizens' views, suggesting areas of divergence and their underlying causes.

In response to these divergences, the need to better incorporate citizens in the governance of secondary data uses is emphasised. Pursuing the law's legitimacy is posited as one requiring ongoing interactions between those using the law and citizens to continually develop and evolve the law over time.

ACKNOWLEDGEMENTS

I am heavily indebted to countless people for the help and kindness during my DPhil studies. I am deeply appreciative of the support provided by Professors Jane Kaye and Bettina Lange over the past few years, and it has been a privilege to take my first tentative steps under their expert guidance.

I am also grateful to those based at the Centre for Health, Law and Emerging Technologies for their unwavering encouragement and particularly the eternal optimism provided by Imogen Holbrook. This research has also benefitted from the influence of the Centre of Socio-Legal Studies, the support of the Economic and Social Research Council, and my time as an intern at Policy Lab.

Beyond the academic sphere, the unwavering support of many friends has been critical to completing this thesis. The combined efforts of Matylda Maciejewska, Tanvi Ravel Mehta, Mao Isobe, Dr Cressida Auckland, and Dr Mohsen Rafii have kept me going through some very difficult times. Finally, my family have, and always will be, my ultimate blessing. Mum, Dad, Nanna, and Grandad – this thesis is dedicated to you, even if you have no idea what I'm going on about.

TABLE OF CONTENTS

Abstract 2

Acknowledgements 4

Table of Contents 5

List of Abbreviations 13

Table of Cases 14

Table of Statutes and Legal Sources 15

List of Tables and Figures 16

I. Introduction 17

1.1 Introducing the Reuse of Administrative Data in Social Science Research 17

1.1.1 Definitions 19

1.1.2 The Opportunities 21

1.1.3 The Challenges and Risks 27

1.1.4 An Overview of the Legislative Responses 31

1.1.5 The Role and Relevance of Societal Expectations 34

1.2 The Approach of this Thesis 37

1.2.1 Research Questions 41

1.3 Thesis outline 42

1.3.1 Part I 43

1.3.2 Part II 46

1.3.3 Part III 50

2 Exploring a Legal Construction of the Permissibility of the Reuse of Data in Research: The General Data Protection Regulation 54

2.1 Introduction 54

2.1.1 Chapter Outline 58

2.2 Generating the Need for the General Data Protection Regulation 59

2.2.1 Competing Constructions of the Role of Rights 61

2.2.2 The Role of Innovation in Constructing Economic and Societal Interests. 66

2.3 Developing the General Data Protection Regulation 71

2.3.1 The Role of Lobbyists 71

2.3.2 The Impact of External Global Events 73

2.3.3 The Impact of the Institutions Within the EU 79

3 Exploring the Effects of the General Data Protection Regulation 95

3.1 Introduction 95

3.2 Silences that Promote Reuses of Data in Research 96

3.3 Discursive Effects that Promote Reuses of Data in Research 101

3.3.1 The 'Truth' is that Developments Provide an Opportunity that Must be Seized. 102

3.3.2 The 'Truth' is that Research Warrants a More Permissive Approach 104

3.3.3 The 'Truth' is that Individual Rights Provide Sufficient Protections 116

3.3.4 The 'Truth' is That Existing Principles Still Fulfil Their Function Within Data Protection Law 119

3.4 Subjectification Effects 122

3.4.1 The Data Subject is Informed and Able to Exercise Control 123

3.4.2 The Data Subject is Supportive of Research. 127

3.4.3 The Data Subject as a Rights-holder in an Improved Position 128

3.5 Lived Effects 130

3.6 Conclusions 132

4 Exploring a Legal Construction of the Permissibility of the Reuse of Data in Research: the Digital Economy Act 2017 133

4.1 Background 133

4.2 Generating the Need for the Digital Economy Act 2017 135

4.2.1 Not Letting the UK 'Fall Behind' 136

4.2.2 Ensuring the Law Meets Societal Expectations. 137

4.2.3 Negative Impacts on Economic and Societal Interests 139

4.3 Developments 141

4.3.1 The Open Policy-Making Process 143

4.3.2 The Better Use of Data in Government Consultation 146

4.3.3 The Impact Assessment 149

4.3.4 The Legislative Process 151
4.4 Effects 155
4.4.1 Silences that Promote Reuses of Data in Research 155
4.4.2 Discursive Effects Construct Research as an Activity that is Permissible Where it Suits Governmental Aims 161
4.4.3 The Effects Lead to a Construction of Adequate Safeguards That Facilitates Data Reuse 168
4.4.4 Subjectification Effects 171
4.5 Concluding Remarks 173
5 An Interlude: Connecting the Doctrinal and the Empirical Work 176
5.1 Introduction 176
5.2 Building Bridges Between the Law and Societal Expectations: The Doctrinal Underpinnings 177
5.2.1 Creating Meaningful Space for Public Participation 179
5.2.2 The Importance of Legitimacy 182
5.2.3 Why Societal Expectations Matter 184
5.3 Building Bridges Between the Law and Societal Expectations: The Empirical Approach 196
5.3.1 Exploring Societal Expectations Through a Survey 200
5.3.2 Exploring Societal Expectations Through Focus Groups 202

6 Investigating Respondents' Views Towards the Acceptability of Secondary Uses of Administrative Data in Research through a Survey 206

6.1 Introduction 206

6.1.1 Methods 210

6.2 Results 211

6.2.1 The Acceptability of Reusing Different Administrative Data Types in Research. 211

6.2.2 The Acceptability of Reusing Data that are More or Less Identifiable in Research 215

6.2.3 The Acceptability of Different Actors Reusing Administrative Data Types in Research 217

6.2.4 The Acceptability of Linking Different Administrative Data Types in Research 219

6.2.5 The Acceptability of Different Purposes of Reusing Administrative Data in Research 221

6.3 The Construction of Research as an Activity in the Public Interest 223

6.4 The Construction of Researchers as Acceptable Actors to Reuse Administrative Data in Research 228

6.5 The Acceptability of Reusing Data that can Identify Individuals 234

6.6 The Acceptability of Reusing Sensitive Data 238

6.7 The Acceptability of Data Linkage 240

6.8 The Ability to Meaningfully Speak of 'Societal Expectations' 247

6.9 Discussion 256

6.9.1 Comparing the Findings to the Existing Literature 257

6.9.2 Taking the Findings Forward: Acknowledging Limitations and the Need for Further Research 267

6.10 Conclusion 270

7 Qualitative Findings I: Exploring the Problematisations Arising from the Focus Group Discussions 271

7.1 Introduction 271

7.2 Justifying the Use of Online Focus Groups 273

7.2.1 The Analytical Approach 274

7.3 Participants' Problematisation of the Reuse of Administrative Data in Research 276

7.3.1 Opportunities Tempered by Risks 276

7.3.2 The Dark Side of Innovations 282

7.4 The Assumptions Underlying the Participants' Problematisations 286

7.4.1 The Trustworthiness of Different Actors 287

7.4.2 Regarding the Trustworthiness of Research 297

7.4.3 Regarding the Concept of Data Ownership 302

7.5 The Factors Prompting the Participants' Problematisations 304

7.5.1 External Events 304

7.5.2 The Market Framing 308

7.6 Reflections and Next Steps 311

8 Qualitative Findings II: The Effects Generated from the Focus Group Discussion 313

8.1 Introduction 313

8.2 The Discursive Effects Generated within the Focus Group Discussions 315

8.2.1 That Individual Control Should Play a Greater Role 315

8.2.2 The Need for Anonymity 323

8.2.3 Questioning Research as an Activity in the Public Interest 329

8.2.4 Legality Fails to Reassure…. 335

8.2.5 …Therefore, Something Needs to Change 338

8.3 The Subjectification Effects Generated within the Focus Group Discussions 343

8.3.1 The Citizen as Overwhelmed and Resigned 343

8.3.2 Actors as Not Necessarily Acting in the Public Interest 347

8.4 Conclusions 350

9 So, What Can We Do? 354

9.1 Introduction 354

9.2 Constructing the Law 356

9.3 Operationalising the Law 367

9.3.1 Lay Participation in Approvals Boards 373

9.3.2 Citizen Juries and Citizen Assemblies 374
9.3.3 vTaiwan 380
9.3.4 Public Data Trusts 389
9.4 Concluding Thoughts and Reflections 391
9.4.1 Future Directions 392
Annex A 395
Mixed Methods 395
Planning 397
The Quantitative Methodology 399
Survey Design 399
Survey Reliability and Validity 402
Survey Recruitment 402
Respondent Demographics 403
Analysis 406
The Qualitative Methodology 406
Justifying the Use of Online Focus Groups. 406
Planning 407
Recruitment 409
The Set-Up 410
The Questioning Route 413
The Focus Groups 416

Conclusion 421

Annex B: Survey 422

Annex C: Focus Group Questioning Route 436

Introduction 436

Provocations 436

Open-Ended Questions 438

Bibliography 440

LIST OF ABBREVIATIONS

Abbreviation	Full Text
ADR UK	Administrative Data Research UK
CUREC	Central University Research Ethics Committee

LIBE	Committee on Civil Liberties, Justice and Home Affairs
DEA	Digital Economy Act 2017
ECHR	European Convention on Human Rights
EDRi	European Digital Rights
EU	European Union
GDPR	General Data Protection Regulation
HMRC	Her Majesty's Revenue and Customs
ICO	Information Commissioner's Office
OECD	Organisation for Economic Co-operation and Development
OPM process	Open Policy Making process
MeCCSA	Media, Communication and Cultural Studies Association
MEPs	Members of European Parliament
PXFGY	Participant X from Focus Group Y
TFEU	Treaty on the Functioning of the European Union

TABLE OF CASES

Case	Page Referenced

Case C-362/14 Maximillian Schrems v Data Protection Commissioner [2015] ECR I-650	74
Joined Cases C-293/12 and C-594/12 Digital Rights Ireland [2014]	116, 125

TABLE OF STATUTES AND LEGAL SOURCES

Statute or Legal Source	Page Referenced
List of UK Legislation	
Data Protection Act 2018	44
Digital Economy Act 2017	33-36, 41, 43-44, 46, 52, 134-144, 146-148, 150-151, 156-158, 161-175, 185-186, 200, 210, 216, 218, 224, 226-229, 233, 235-236, 238-239, 241, 248, 257-258, 260, 264-266, 268, 271-273, 276, 279-286, 288, 291, 293-294, 296-298, 300-301, 303-305, 308-313, 315-316, 318, 324-328, 330-332, 335, 338, 334, 348-349, 352, 357, 368, 374, 393, 395, 416, 420-421
List of EU Legislation	
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 (Data Protection Directive)	32, 68, 97, 123
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	33-36, 41, 43-45, 52, 55, 57-63, 67-68, 72-77, 79-80, 83, 87-89, 91-93, 95-109, 111-114, 116-117, 121-127, 129-131, 133-134, 139, 185-186, 200, 210, 216, 218, 224, 226, 228-229, 233, 235-242,

such data, and repealing Directive 95/46/EC (General Data Protection Regulation)	247-248, 257-258, 260, 264-266, 268, 271-273, 276, 279-286, 288-289, 291, 298-301, 303-305, 308-313, 315-316, 318, 323-328, 330-332, 335, 338, 344-345, 347-349, 352, 357, 368, 393, 395, 400, 408, 416
Other Legal Sources	
European Convention on Human Rights	90, 108

LIST OF TABLES AND FIGURES

Tables and Figures	Page Provided
List of Tables	
Table 1	213
Table 2	215
Table 3	217
Table 4	219
Table 5	221
Table 6	223
Table 7	244
Table 8	246
Table 9	250
Table 10	254
Table A.1	405
Table A.2	405
Table A.3	406
Table A.4	406

Table A.5 412

Table A.6 413

Table A.7 413

List of Figures

Figure 1 224

Figure 2 225

Figure 3 226

Figure 4 227

Figure 5 231

Figure 6 231

Figure 7 233

Figure 8 237

I. INTRODUCTION

1.1 INTRODUCING THE REUSE OF ADMINISTRATIVE DATA IN SOCIAL SCIENCE RESEARCH

In 2013, the UK Government identified 'big data' as a technology that would contribute to the future of UK growth, allocating £189 million to its development.[1] The same year saw the Department for Business, Innovation and Skills herald administrative data as a 'unique resource' with 'the potential to develop new

[1] Science and Technology Committee, 'The Big Data Dilemma' HC (2015-16) 468 [3] <https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/46802.htm>.

understanding and insights between different fields of study and over time.'[2] Reusing administrative data is perceived as 'crucial' for developing policies and services that are sustainable, inclusive, and trusted in the long term.[3] Established in 2018, the Administrative Data Research UK (ADR UK) built on previous initiatives[4] to 'maximise the potential of administrative data as a resource for high-quality research' to create 'impact around prioritised policy themes within the UK.'[5] This initiative has fostered ground-breaking research in the context of crime and justice,[6] education,[7] and homelessness.[8]

These developments raise several broader questions. When data are collected for one purpose, what could justify its reuse and repurposing in a different context? If justifications are rooted in notions of 'the public interest' and the ability to secure certain social goods, what are the implications of intra-societal disagreement with specific data processing activities being categorised this way?

---

[2] Department for Business Innovation & Skills, 'Improving Access for Research and Policy: The Government Response to the Report of the Administrative Data Taskforce' (2013) <https://www.gov.uk/government/publications/administrative-data-taskforce-report-government-response>.

[3] Charlotte Van Ooijen, Barbara Ubaldi and Benjamin Welby, 'A Data-Driven Public Sector: Enabling the Strategic Use of Data for Productive, Inclusive and Trustworthy Governance' [2019] <https://doi.org/10.1787/19934351>.

[4] Economic and Social Research Council, 'Big Data Network' (2018) <https://esrc.ukri.org/research/our-research/big-data-network/>.

[5] ADR UK, 'About ADR UK' (2018) <https://www.adruk.org/about-us/about-adr-uk/>.

[6] ADR UK, 'Data First: Harnessing the Potential of Linked Administrative Data for the Justice System' (2020) <https://www.adruk.org/our-work/browse-all-projects/data-first-harnessing-the-potential-of-linked-administrative-data-for-the-justice-system-169/>.

[7] ADR UK, 'Understanding Children's Lives and Outcomes in Scotland' (2020) <https://www.adruk.org/our-work/browse-all-projects/understanding-childrens-lives-and-outcomes-in-scotland-140/>.

[8] ADR UK, 'Homelessness Data England: Linking Local Authority Data to Evaluate Homelessness Policy' (2020) <https://www.adruk.org/our-work/browse-all-projects/homelessness-data-england-linking-local-authority-data-to-evaluate-homelessness-policy-267/>.

The law refers to individuals' reasonable expectations towards using their data as relevant to determining which activities should be legally permissible. What happens when, due to a widespread unfamiliarity with modern data processing activities, it becomes nearly impossible to speak of 'expectations' meaningfully.

This thesis explores these questions in the context of secondary uses of administrative data in social science research. This section introduces and defines this data processing activity before justifying the scope of the thesis. It explores the new opportunities and challenges of applying innovative data processing practices, collectively known as 'big data analytics,' to administrative data. Big data analytics can generate novel insights that may be useful to policymakers in government, commercial entities, and other sectors of society. The section provides an overview of societal and legislative responses to these developments and their importance to the thesis. The rest of the chapter provides an overview of the current landscape to justify the focus on the secondary uses of administrative data in social science research, explains how the landscape informed the research questions and the research approach, and concludes with the thesis outline.

1.1.1 Definitions

This research focuses on secondary uses of administrative data in the UK. This encompasses a broad range of information collected by the UK government on its citizens. The literature lacks a unified definition of 'administrative data,' so this thesis uses Woollard's definition to classify it as

> *information collected primarily for administrative purposes. This type of data is collected by government departments and other*

organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.[9]

Registration entails recording specific information for reference purposes, including the registration of births and deaths and the electoral register. Transactional data capture information for the government to fulfil its responsibilities, including information to inform tax payments and the provision of certain benefits.[10] Sources of administrative data in the UK include the Department for Work and Pensions Customer Information Systems,[11] the Police National Computer,[12] the English and Welsh School Censuses,[13] and the Electoral Register. 'Secondary uses' refers to uses of this data beyond its original purpose. The focus on secondary uses arises from the legislative approach towards governing the processing of data for research purposes, which treats this reuse of data for this purpose in a unique way. By examining social science research specifically, this thesis explores research that has attracted little attention in the existing literature that has considered the legal approach and societal views towards reuses of data in research. It is acknowledged from the outset that it may

[9] Matthew Woollard, 'Administrative Data: Problems and Benefits. A Perspective from the United Kingdom' in Adrian Duşa and others (eds), *Facing the Future: European Research Infrastructure for Humanities and Social Sciences* (SCIVERO 2014).

[10] Ibid.

[11] Department for Work & Pensions, 'Customer Information System (CIS): The Information Held about You' (2018) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/729893/customer-information-system-the-data-held-about-you.pdf>.

[12] Home Office, 'Guidance – Police National Computer (PNC)' (2014) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/873384/PNC_v5.0_EXT_clean.pdf>.

[13] Department for Education, 'Complete the School Census' (2019). <https://www.gov.uk/guidance/complete-the-school-census>.

be impossible to define 'social science research' to delineate it fully from other types of research, such as medical research. The thesis uses the definition proposed by the ESRC as research that

> *tells us about the world beyond our immediate experience, and can help explain how our own society works - from the causes of unemployment or what helps economic growth, to how and why people vote, or what makes people happy. It provides vital information for governments and policymakers, local authorities, non-governmental organisations and others.*[14]

This definition helps to elucidate the potential relevance of social science research to governments, which may better understand societal problems as a result. It shows the prominent role of citizens within the social sciences, which considers citizens' experiences and explores why things happen to them. Finally, it indicates the potential for a fruitful relationship between researchers and the government.

1.1.2 The Opportunities

A combination of technological developments has led to the growth of 'big data analytics,' enabling innovative leaps in data-driven analysis and decision-making. These technological developments have allowed researchers to begin unlocking the potential of administrative data to provide novel insights into modern societal challenges. A detailed exploration of this phenomenon is unnecessary to address the questions explored within this thesis. Instead, this section will provide an overview of the potential and application of big data analytical approaches that

[14] Economic and Social Research Council, 'What Is Social Science?' (2018) <https://esrc.ukri.org/about-us/what-is-social-science/>.

may drive new legislation and government policies, giving the context and background necessary to justify the focus of this thesis.

The first set of developments relates to innovations in technology, which have led to substantial improvements in the ability to collect, store, and link ever-increasing amounts of data. Many consider the ability to accumulate a vast volume of data as one of the 'hallmarks' of 'big data.'[15] Further characteristics commonly associated with big data are its variety, which encompasses the wide range of forms data can take; its velocity, which focuses upon the speed with which data are created, stored, and retrieved; and its complexity, which considers the degree to which the data are interconnected.[16] The granularity, complexity, and exhaustive scope of administrative data have led to its recognition as an example of big data.[17]

Alongside these advances, new analytical techniques, including machine learning and data mining, capitalise on the greater amounts of data available.[18] Machine learning is an analytical technique that seeks to 'iteratively evolve an

[15] Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences* (SAGE Publications, 2017); Fola Malomo and Vania Sena, 'Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector' (2016) 9 Policy & Internet 7; Yvonne McDermott, 'Conceptualising the Right to Data Protection in an Era of Big Data' (2017) 4 Big Data & Society 1.

[16] Kevin C Desouza and Benoy Jacob, 'Big Data in the Public Sector: Lessons for Practitioners and Scholars' (2014) 49 Administration & Society 1043.

[17] Roxanne Connelly and others, 'The Role of Administrative Data in the Big Data Revolution in Social Science Research' (2016) 59 Social Science Research 1; Justin Longo and Alan Rodney Dobell, 'The Limits of Policy Analytics: Early Examples and the Emerging Boundary of Possibilities' (2018) 6 Politics and Governance 5.

[18] Irina Pencheva, Marc Esteve and Slava Jankin Mikhaylov, 'Big Data and AI – a Transformational Shift for Government: So, What Next for Research?' (2018) 35 Public Policy and Administration 24; Chris Yiu, *The Big Data Opportunity: Making Government Faster, Smarter and More Personal* (Policy Exchange 2012).

understanding of a dataset'[19] to construct increasingly sophisticated models that explain identified patterns and predict future outcomes. Data mining enables data extraction and identifying patterns from large datasets based on the premise that these huge datasets hold meaningful information.[20] Combining these tools and techniques has fostered 'a new way of thinking about knowledge through data and a framework for supporting decision making, rationalizing action, and guiding practice.'[21]

'Big data analytics,' therefore, can be understood as a more powerful version of knowledge discovery than what was previously possible. This 'enables the inference of previously unknown links on patterns, relying on finding correlations rather than causation.'[22] Governments are increasingly extolling the virtues of opening up their administrative data to third parties for secondary uses in recognition of the potential benefits of using big data analytics. Included within these 'virtues' are research activities reusing administrative data to better understand and respond to modern societal problems.[23] The HMRC Datalab is an example of the UK government attempting to open up its administrative data to

[19] Kitchin (n 15).

[20] Ibid.

[21] Solon Barocas and Helen Nissenbaum, 'Big Data's End Run around Anonymity and Consent' in Julia Lane and others (eds), *Privacy, Big Data and the Public Good: Frameworks for Engagement* (Cambridge University Press 2014).

[22] Ira Rubinstein, 'Big Data: The End of Privacy or a New Beginning?' (2013) 3 International Data Privacy Law 74.

[23] Cabinet Office, 'Better Use of Data in Government' (*GOV.UK,* February 29, 2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government>; Big Data Senior Steering Group, 'The Federal Big Data Research and Development Strategic Plan' (National Science and Technology Council 2016) <https://www.nitrd.gov/pubs/bigdatardstrategicplan.pdf>.

researchers where such research is in the 'public interest.'[24] Through this initiative, academic researchers may access large administrative datasets on Corporation Tax and VAT Returns, Trade Statistics, and Stamp Duty Land Tax.[25] Such developments are not unique to the UK. Countries including the USA,[26] Canada,[27] and Australia[28] have all taken steps to ensure their administrative data is more widely available to researchers.

Governments frequently stress the benefits of researchers using big data analytics on administrative data. Using big data analytics permits an 'unprecedented comprehensiveness'[29] to understand better the current policy landscape, improve the identification of problems, and determine the most optimal solutions.[30] The enhanced predictive capabilities of big data analytics also respond to a key challenge of traditional policymaking, whereby policy implementation

[24] Richard Welpton and Melanie Wright, 'HMRC Datalab: Engaging with the External Research Community' (2011) <https://dam.data-archive.ac.uk/reports/research/hmrcdatalab.pdf>.

[25] HM Revenue & Customs, 'Open Data Strategy' (2012) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/89258/open-data-strat.pdf>.

[26] David E Card and others, 'Expanding Access to Administrative Data for Research in the United States' [2010] SSRN Electronic Journal <https://dx.doi.org/10.2139/ssrn.1888586>.

[27] R Finnie, M Dubois, and M Miyairi, 'How Student Pathways Affect Labour Market Outcomes: Evidence from Tax-Linked Administrative Data' (2017) <https://www.oncat.ca/sites/default/files/research-project-attachments/2016-08-final-report-university-of-ottawa-how-student-pathways-affect-labour-market-outcomes.pdf>.

[28] David Smith and Therese Smith, 'Using Linking to Sharpen Policy-Thinking: Early Findings from the Australian Census and Migrants Integrated Dataset' (2014) 47 Australian Economic Review 220.

[29] Viktor Mayer-Schönberger and Yann Padova, 'Regime Change? Enabling Big Data through Europe's New Data Protection Regulation' (2016) 17(2) Science and Technology Law Review <https://doi.org/10.7916/stlr.v17i2.4007>.

[30] Pencheva and others (n 18).

often leads to unintended and unwanted consequences.[31] Big data analytics ameliorates these challenges by providing a deeper understanding of the relations and dependencies in existing datasets to facilitate superior decision-making processes.[32] Some see the development of a data-driven public sector as offering 'huge opportunities' for policymaking and service provision[33] to provide wide public benefits and social outcomes that 'enhance the wellbeing of people and communities.'[34] Thus, discussions are often rooted in the idea that using big data analytics is in the 'public interest' by providing 'tailormade, context-specific information and knowledge informed by an analysis of data.'[35]

The research community also shares an interest in using administrative data this way. Whilst not without its own challenges,[36] the potential of big data analytics mean researchers are keen to access and analyse big data, including government-held administrative data. Big data analytics may hold 'great potential' for how researchers can understand society, bringing a 'major shift in the way that social science is conducted.'[37] By 2020, the ADR UK had supported

[31] Ibid.

[32] José María Cavanillas, Edward Curry and Wolfgang Wahlster (eds), *New Horizons for a Data-Driven Economy* (Cham Springer International Publishing, 2016).

[33] Cat Drew, 'Data Science Ethics in Government' (2016) 374 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20160119.

[34] Involve, 'Data for Public Benefit: Balancing the Risks and Benefits of Data Sharing' (2018) <https://www.involve.org.uk/resources/publications/project-reports/data-public-benefit-balancing-risks-and-benefits-data-sharing>.

[35] Van Ooijen (n 3).

[36] Connelly and others (n 17).

[37] Amanda Clarke and Helen Margetts, 'Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform' (2014) 6 Policy & Internet 393.

nearly 1000 research projects and awarded over £5 million of grants to support 'innovative research of importance to key public policy areas.'[38] The reductions in time and costs associated with researchers capitalising upon the developments in big data analytics give them the ability to open up 'a new perspective on reality.'[39] This analysis of administrative data could provide new insights into social inequality, human behaviours, and the effectiveness of social policies.[40] The comprehensive nature of big data reduces some of the possibilities of bias error.[41] The significant increase in sample size means data may cover entire populations, which can assist researchers where their interest lies in small sub-groups, rare events, or those usually unwilling or unable to participate in research.[42] Furthermore, researchers are under increasing pressure to demonstrate their research has policy impact, most notably through the Research Excellence Framework case studies[43] and the 'Pathways to Impact' developed by UK Research Councils. The purpose of these initiatives is to encourage researchers to consider ways 'to ensure [their] research makes a difference.'[44]

---

[38] ADR UK, 'ADR UK Annual Report 2019–20' (2020) <https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADR_UK_annual_report_2019_20.pdf>.

[39] Mayer-Schönberger and Padova (n 29).

[40] Connelly and others (n 17).

[41] Mayer-Schönberger and Padova (n 29).

[42] Connelly and others (n 17).

[43] Research Excellence Framework, 'Index of Revisions to the 'Panel Criteria and Working Methods' (2019/02)' (2020) <https://www.ref.ac.uk/media/1084/ref-2019_02-panel-criteria-and-working-methods.pdf>.

[44] UK Research and Innovation, 'Pathways to Impact' (2018) <https://webarchive.nationalarchives.gov.uk/20200203113953/https://www.ukri.org/innovation/excellence-with-impact/pathways-to-impact/>.

1.1.3 The Challenges and Risks

The framing of the use of big data analytics in the previous section aligns with Kitchin's argument that such developments are not unfolding in a non-ideological way. Instead, a 'powerful set of arguments, forwarded by passionate believers in the benefits of new ways of knowing and acting in the world'[45] drive these developments. Proponents often reference data analytics being founded upon 'hard' scientific evidence to legitimise data-driven policymaking.[46] Beneath the claims of a more effective and efficient service delivery method are assumptions concerning the level of knowledge and objectivity fostered by big data analytics.[47] However, using scientific knowledge in policy does not entail a linear process from science to policy.[48] Instead, such engagement is fluid, and the interactions are co-productions between the scientisation of politics and the politicisation of science. Apolitical perceptions of scientific knowledge thus enable the generation of political decisions, which can obscure the political character of the decisions made.[49]

Whilst governments may focus on the potential benefits from using big data analytics for wider society, often based on notions of the public interest, these developments are not universally celebrated. The potential benefits of policy

[45] Kitchin (n 15).
[46] Gernot Rieder and Judith Simon, 'Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data' (2016) 3 Big Data & Society.

[47] Ibid.

[48] Anna Wesselink and others, 'Technical Knowledge, Discursive Spaces and Politics at the Science–Policy Interface' (2013) 30 Environmental Science & Policy 1.

[49] Ibid.

driven by big data analytics do not mean that it should be perceived as a 'silver bullet' solution for difficult policymaking decisions. Big data analytics do not preclude misguided and mistaken uses of data.[50] Furthermore, some note that scientific findings do not automatically translate into better policy, and evidence may be susceptible to conflicting politically driven priorities.[51] The OECD recognises that having the '"right" information at the "right" time does not automatically lead to the "right" decision.' Instead, this concept depends both on the decision made and the process leading up to it.[52] As Weiss notes, findings may be used to support pre-existing policy preferences or to delay the decision-making process to provide the policymaker with breathing space.[53] Boswell and Smith argue that efforts to bring researchers and policymakers closer together may not be a desirable outcome as it risks the increasing politicisation of research and an associated reduction in academic freedom.[54]

Big data analytics are not premised upon the traditional research method of testing a hypothesis with available data. Instead, sophisticated algorithms are

[50] Katherine J Strandburg, 'Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context' in Julia Lane and others (eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Cambridge University Press 2014).

[51] Johann Höchtl, Peter Parycek and Ralph Schöllhammer, 'Big Data in the Policy Cycle: Policy Decision Making in the Digital Era' (2015) 26 Journal of Organizational Computing and Electronic Commerce 147.

[52] Van Ooijen (n 3).

[53] Janet A Weiss, 'Access to Influence' (1979) 22 American Behavioral Scientist 437.

[54] Christina Boswell and Katherine Smith, 'Rethinking Policy 'Impact': Four Models of Research-Policy Relations' (2017) 3 Palgrave Communications.

deployed on datasets, fostering an inherent unpredictability in their design.[55] This ability to obtain hidden inferences and correlations makes the results and the decision-making rationale of these algorithms increasingly opaque to humans.[56] The value of big data arises from its ability to be repurposed as governments need not spend additional money collecting, storing, and analysing data under their control. However, sharing data with third parties and repurposing data increases the distance between its source and eventual use. Doing so increases the gap between individuals' expectations when the data were first collected, and their various uses, raising questions of responsibility.[57]

Big data analytics may also engender new and greater intrusions into individuals' privacy interests. The mandates of the public agencies of governments mean that the vast amounts of information gathered on a citizen could elucidate 'an almost complete picture' of their life.[58] The linkage of increasing amounts of data increases the risk of 'deductive disclosure,' whereby de-identified data is rendered re-identifiable. Big data analytics weaken the effectiveness of anonymisation, a key technique for protecting individuals' privacy interests.[59] As big data analytics are premised upon the reuse of data previously collected for

[55] Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' [2018] ico.org.uk <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

[56] Ibid.

[57] M Steinmann, SA Matei and J Collmann, 'A Theoretical Framework for Ethical Reflection in Big Data Research' in SA Matei and J Collmann (eds), *Ethical Reasoning in Big Data: An Exploratory Analysis* (Springer Netherlands 2016).

[58] Pencheva and others (n 18).

[59] Rubinstein (n 22).

administrative purposes, it undermines individuals' ability to control their data and consent to its use. Traditionally, the mechanisms of anonymity and consent have operated to delineate the boundaries between acceptable and unacceptable data processing activities.[60] This use of data, which circumnavigates these mechanisms, may upset the balance of social and institutional power. There is significantly more scope for data not to mirror an individual's circumstances and exclude such individuals from decisions regarding the uses of their information.[61]

These risks are not purely theoretical. Numerous examples of government-driven initiatives that sought to capitalise upon data analytics have prompted controversy, demonstrating their real-world consequences. In the UK, the poor-quality algorithm that incorrectly found many international students to have cheated on their English proficiency tests caused a significant uproar, with many students successfully challenging these decisions.[62] In Australia, there was significant controversy over the introduction of an algorithm used to recover overpayments from welfare recipients because 15% of detected discrepancies were not debts owed.[63] Furthermore, Illinois abandoned its attempts to use data mining

[60] Alan Westin, *Privacy and Freedom* (Ig Publishing 2015).

[61] Marion Oswald, 'Share and Share Alike? An Examination of Trust, Anonymisation and Data Sharing with Particular Reference to an Exploratory Research Project Investigating Attitudes to Sharing Personal Data with the Public Sector' (2014) 11(3) SCRIPTed 245.

[62] BBC News, 'Students 'May Have Been Unfairly Deported' over English Test Cheat Claims' *BBC News* (May 24, 2019) <https://www.bbc.co.uk/news/uk-48394001>.

[63] Paul Karp, 'Government Knew about Discrepancies in Data-Matching System before Reducing Human Oversight' *the Guardian* (January 11, 2017) <https://www.theguardian.com/australia-news/2017/jan/12/centrelink-human-oversight-slashed-after-cost-benefit-analysis>.

to identify at-risk children due to the 'unreliable'[64] technology. It is, therefore, unsurprising that lawmakers have recognised such developments. The following section will introduce some recent legislative instruments that help govern the reuse of data in research explored in Part I of the thesis.

1.1.4 An Overview of the Legislative Responses

The thesis focuses on two legislative instruments that govern data processing activities. The Janus-faced nature of big data analytics, which can potentially foster significant benefits and significant harms to society, is recognised as a 'conundrum' that the law must face.[65] As Strandburg notes, there is a need to 'provide space for big data to fulfil its potential for societal benefit, while protecting citizens adequately from related individual and social harms.'[66]

In 2012, the EU Commission recognised the 'rapid pace of technological change' had 'profoundly changed' how personal data are collected, accessed, used, and transformed.[67] At that time, the central legislation governing personal data protection in the EU was the Data Protection Directive,[68] which entered into force in 1995. The Commission argued that its twin objectives of ensuring a functioning

[64] Gary Marx and David Jackson, 'Data Mining Program Designed to Predict Child Abuse Proves Unreliable, DCFS Says' *chicagotribune.com* (December 6, 2017) <https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html>.

[65] Strandburg (n 50).

[66] Ibid.

[67] Commission, 'Safeguarding Privacy in a Connected World a European Data Protection Framework for the 21st Century (Text with EEA Relevance)' (Communication) COM (2012) 9 final, 25 January 2012.

[68] 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.

Single Market and the effective protection of individuals' rights were still valid. However, the existing rules were inadequate to promote them.[69] To enhance individuals' rights, the Commission argued in favour of reforms to improve individuals' abilities to control their data, exercise their rights, and enhance the accountability of those processing data.[70] This culminated in the introduction of the General Data Protection Regulation[71] (GDPR) by the EU in May 2016. The Digital Economy Act 2017[72] (DEA) received Royal Assent in April 2017.

It implemented several policies outlined in the 2015 Conservative Party Manifesto, including the 'delivery of better public services, world-leading research and better statistics.'[73] Part 5 concerns 'Digital Government', and Chapter 5 of Part 5 introduces a specific legal gateway for researchers to use data initially collected for administrative purposes.

Both legislative instruments refer to societal expectations and norms in justifying the governances of reusing data in research. The GDPR references explicitly the 'expectations'[74] of individuals to help guide data controllers' and data processors' determination of the acceptability of a data processing activity. Research is constructed as a means of obtaining 'essential knowledge about the

[69] Commission (n 67).

[70] Ibid.

[71] EU General Data Protection Regulation (GDPR): 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.

[72] Digital Economy Act 2017.

[73] Explanatory Notes to the Digital Economy Act 2017.

[74] See Recital 50 GDPR.

long term correlation of a number of social conditions, such as unemployment and education with other life conditions.'[75] Research is further portrayed as providing 'solid, high-quality knowledge which can provide the basis for the formulation and implementation of knowledge-based policy' to help improve people's life quality and the efficiency of social services.[76] As a result, data processing for research purposes benefits from the presumption that further processing for research purposes is compatible with the initial purpose of collection, facilitating this reuse of data under Article 5(1.) Data processing for research purposes may benefit from several derogations and limitations from the requirements of Articles 85 and 89. The GDPR also constructs the right to the protection of personal data as one to be considered 'in relation to its function in society.'[77]

The DEA permits only de-identified data to be shared with researchers, which is defined as data that do not directly identify individuals and are not reasonably likely to lead to an individual's identity being ascertained. Societal expectations are highly relevant as the legislation requires research to be accredited as 'in the public interest' before it can benefit from access to administrative data.[78]

[75] See Recital 157 GDPR.

[76] Ibid.

[77] See Recital 4 GDPR.

[78] See s71 Digital Economy Act 2017.

1.1.5 The Role and Relevance of Societal Expectations

Societal responses are important to this thesis in two ways. Firstly, the legislation governing the secondary uses of data recognises that data processing can contribute to the prosperity and welfare of society, yet it is not an activity that can be undertaken unrestrained. Underlying certain GDPR provisions is an assumption that, due to the social goods certain data processing activities can promote, they are justified as activities in the 'public interest.' [79] Reusing administrative data in social research may be categorised this way due to the insights into 'social human behaviours and the effectiveness of social policies'[80] it can offer governments when they make and reform their policies.[81] The GDPR's Recitals highlight and emphasise the contributions research can make. The DEA's approach is even more explicit, using the 'public interest' as a criterion that a research project must demonstrate before administrative data can be shared.[82]

Secondly, there is a persistent issue regarding the uncertainty surrounding what individuals' rights protect and what may be considered as a data processing activity in the public interest. In the complex and ever-evolving data processing landscape that operates outside the awareness of most individuals in society, how are such terms to guide what is acceptable or expected? This uncertainty risks the interpretation of these terms being biased towards the objectives and beliefs of data controllers and data processors in their determinations.

[79] See Recital 51 and Article 9, EU General Data Protection Regulation (GDPR).

[80] Connelly and others (n 17).

[81] See Articles 21 and 55 EU General Data Protection Regulation (GDPR).

[82] See s71(4) Digital Economy Act 2017.

Ambiguity as to what should be protected, what should be promoted, and how different interests should be reconciled creates scope for uncertainty and disagreement. Whilst a data processing activity may be legally permissible, if it does not align with how individuals conceive their interests, there is scope for such activities to be hindered regardless of their wider potential public benefit. Care.data demonstrates that such issues are not theoretical.[83] Care.data was a legally permissible data-sharing initiative that sought to open up government-held medical data for various purposes, including research. The initiative, however, failed to align with individuals' expectations as to how their data would be used and shared.[84] This triggered a severe public backlash and the eventual cancellation of the project.[85] The outcome suggests that legal authority is insufficient alone to secure societal acceptance of data processing activities. It also indicates a need to further understand individuals' attitudes in this area and reflect upon their views on the governance of reusing data in research. These controversies have highlighted the importance of engaging with the public to ensure the data processing activities align with the values and interests those individuals perceive as relevant in this context.[86] The GDPR and DEA employed public consultations during the early stages of the legislative process to allow the

[83] Fiona Godlee, 'What Can We Salvage from Care.data?' (2016) 354 BMJ 3907.
[84] Rebecca Hays and Gavin Daker-White, 'The Care.data Consensus? A Qualitative Analysis of Opinions Expressed on Twitter' (2015) 15 BMC Public Health 838.

[85] Sigrid Sterckx and others, "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' (2015) 19 Medicine, Health Care and Philosophy 177.

[86] Mhairi Aitken and others, 'Who Benefits and How? Public Expectations of Public Benefits from Data-Intensive Health Research' (2018) 5 Big Data & Society.

public to contribute their views, supporting this idea that societal views play a role in legitimising the law.

The use of the 'public interest' and societal benefits to justify reuses of data warrants the further exploration of the societal constructions of these phrases. In contrast to empirical research exploring attitudes towards the reuse of data in social research, there is significant literature exploring individuals' attitudes towards using their health data. Existing studies can provide insights into how the public think about the secondary use of data. However, the studies indicate that findings from research exploring attitudes towards the reuse of health data in medical research are not generalisable to other forms of research, using different kinds of data, such as the administrative data analysed in this thesis.[87]

In contrast to medical research, the empirical research conducted on the public's views on reusing administrative data in research is quite limited, though the literature in this area is growing.[88] This thesis contributes to this field through its socio-legal approach to the issues raised in this chapter, enabling a richer interdisciplinary exploration of the research questions. Previous work has tended to focus on individuals' views[89] or what is legally permissible.[90] A socio-legal approach enables a consideration of both, providing a richer understanding of the issues and novel insights into possible responses to potential divergences between

[87] Daniel Cameron, Sarah Pope and Michael Clemence, 'Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes' (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/>.

[88] Ibid.

[89] Ibid.

[90] Mayer-Schönberger and Padova (n 29).

the law and societal expectations. Furthermore, using a mixed-methods approach allows for the collection of a breadth of views and the deeper exploration of certain views. A combination of qualitative and quantitative methods, through an online survey and a series of 5 online focus groups, help provide a deeper understanding of citizens' views towards the reuse of administrative data in research. The role and relevance of societal expectations, and the methods used to elicit them, will be discussed further in the next section, which outlines the approach of this thesis. Annex A provides a detailed methodology.

1.2 THE APPROACH OF THIS THESIS

This section outlines the thesis' approach and key research questions. Fundamental to this thesis is the premise that societal attitudes and expectations contribute to the quality and legitimacy of law thus should play a role in the legislative process. This draws on ideas of legitimacy defined as the 'generalised perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions.'[91] This understanding of legitimacy goes beyond a law fulfilling a 'pedigree test' of validity through the 'stamp of official authority.'[92] Instead, it aligns with the understanding of 'social legitimacy' proposed by Taylor and Whitton. They draw upon the works of Lipset[93] to find legitimacy to be associated

[91] Mark Suchman, 'Managing Legitimacy: Strategic and Institutional Approaches' (1995) 20 Academy of Management 571, 574.

[92] Philippe Nonet and Philip Selznick, *Law and Society in Transition: Toward Responsive Law* (Transaction Publishers 2017), 83.

[93] Mark Taylor and Tess Whitton, 'Public Interest, Health Research and Data Protection Law: Establishing a Legitimate Trade-off between Individual Control and Research Access to Health Data' (2020) 9 Laws 6.

with a system's capacity to maintain the belief that existing political institutions are the most appropriate for society.[94] In doing so, the exercise of political power is justified where it aligns with the constitutional essentials that all citizens may reasonably be expected to endorse, considering the principles and ideals acceptable to them as reasonable and rational.[95] This understanding is consistent with Poort's assessment of the law's legitimacy being somewhat premised with its fit with the social context, defined as the potential for societal input into developing norms.[96] Poort's writings align with Selznick's claims that, for a well-ordered legal system, 'nothing is more important than social support.'[97] The norms in this context are difficult to determine from the outset due to the complexity of innovations and lack of widespread public awareness. Therefore, this thesis sees societal expectations potentially feeding into the norms present in the law even after its enactment, a point explored in Part III of the thesis.

This thesis, therefore, acts almost as a case study of the law's legitimacy where divergences between the law and societal views arise and, based on this, considers the potential responses to identified divergences. In considering these potential responses, the concept of 'responsive law'[98] is considered. This concept

---

[94] Seymour Martin Lipset, *Political Man: The Social Bases of Politics* (The Johns Hopkins University Press 1994), 64.

[95] Lawrence B Solum, 'Public Legal Reason' (2006) 92 Virginia Law Review 1472.

[96] Lonneke Poort, 'Consensus & Controversies: An Interactive Legislative Approach to Animal Biotechnology in Denmark, Switzerland and the Netherlands' (PhD thesis, Tilburg University 2011) <https://research.tilburguniversity.edu/en/publications/consensus-amp-controversies-an-interactive-legislative-approach-t>.

[97] Philip Selznick, 'Law in Context' Revisited' (2003) 30 Journal of Law and Society 177, 178.

[98] Nonet and Selznick (n 92).

emphasises the importance of an ongoing relationship between the law and citizens. 'Responsive law' is distinguished from repressive law, where the focus is on attaining a ruler's objectives in a top-down fashion, and autonomous law, which emphasises the upholding of a formal understanding of the rule of law that focuses on the centrality of rules. Responsive law goes further than upholding a formalistic understanding of justice to a more substantive one where rules are treated as guides towards promoting and upholding more general values and purposes. Rather than operating as exhaustive reasons for action, rules act as 'grounds of criticism, as rules and decisions come to be evaluated according to underlying purposes.' [99] Here, the law becomes 'an instrument for change…open to reconstruction in the light of how those who are governed perceive their rights and reassess their moral commitments.' [100] Whilst Selznick and Nonet define legitimacy as the key concern of autonomous law, the understanding of legitimacy used in this thesis goes beyond the demonstration of formal requirements that enable the law to meet a 'pedigree test.'[101] Instead, it aligns with the purposive results-oriented approach at the heart of the ideas surrounding responsive law in its consideration of societal norms and values.

This construction of law responds to a potential issue arising from the understanding of legitimacy used in this thesis. The difficulties in defining a single 'socially constructed system of norms, values, beliefs, and definitions'[102] due to the

[99] Martin Krygier, *Philip Selznick: Ideals in the World* (Stanford Law Books 2012), 179.

[100] Nonet and Selznick (n 92).

[101] Nonet and Selznick (n 92), 83.

[102] Suchman (n 91), 574.

diversity of views is acknowledged. In response, legitimacy is not a fixed attribute of the law at its enactment, depending on the ability of the legislative process to produce a law that fits with societal views of what is appropriate based on a shared system of beliefs and values. Instead, the notion of responsive law creates space for legitimacy to be an ongoing pursuit of law through continued interactions with society. In Nonet and Selznick's articulation of responsive law, they highlight as a distinctive feature the search for implicit values in rules and, as these values are articulated, 'they offer authoritative criteria for criticising existing rules' and 'generating new rules.'[103] This approach provides some flexibility within the law that may better respond to the diverse views and beliefs held by those within society. The law's legitimacy is not a fixed attribute to be determined at a fixed point in time but rather a quality that can be further developed in the post-legislative space. This may occur through the recognition that as 'circumstances alter, rules must be refashioned, not only to meet the needs of policy but also to protect the authority of the rules themselves and the integrity of their application.'[104] The work, therefore, ties in with the 'aspiration for a legal order that is effective in dealing with change.'[105] Three key research questions and a socio-legal approach is used in this thesis to explore these points in a structured way. The following sections further detail this approach.

[103] Nonet and Selznick (n 92), 79.

[104] Nonet and Selznick (n 92), 80.

[105] Nonet and Selznick (n 92), 81.

1.2.1 Research Questions

I. How do the GDPR and the DEA 2017 respond to secondary uses of administrative data in social science research?

II. How far do the legislative approaches align with citizens' views towards reusing administrative data in social science research?

III. In light of identified differences, or where a lack of societal consensus undermines attempts to explore the notion of 'alignment', how can legislation and those using legislation to determine appropriate secondary data uses respond?

A socio-legal approach underpins the exploration of these questions. Whilst no unified definition of a 'socio-legal approach' exists, it is understood as an approach that places the law in a broader sociocultural context, enabling its study in relation to societal forces.[106] This acknowledges that unless 'open to the full range of norms in play,'[107] it is a fruitless task to try to understand why certain activities, such as data processing, may be lawful yet provoke a negative societal response. This resonates with Pound's claim that such an approach enables a 'more complete and intelligent account of the social facts upon which the law must process and to which it is to be applied.'[108] Where the law aims to influence actions, categorising activities as permissible and impermissible, a deeper understanding

[106] Reza Banakar, 'Having One's Cake and Eating It: The Paradox of Contextualisation in Socio-Legal Research' (2011) 7 International Journal of Law in Context 487.

[107] Roger Brownsword, 'In the Year 2061: From Law to Technological Management' (2015) 7 Law, Innovation and Technology 1.

[108] Roscoe Pound, *Jurisprudence* (West Pub Co 1959), 350.

of how the law constructs these meanings and categories and how individuals regard them is imperative.[109] The thesis contributes to this understanding by exploring these legal constructions and the societal responses in the context of reusing administrative data in social science research. This exploration follows Nonet and Selznick's approach regarding the preference for exploraing to what extent and under what conditions connections between law and the state, law and rules, and law and moral aspiration occur. [110]

1.3 THESIS OUTLINE

This introduction provided the necessary context for understanding this thesis by defining key terms such as 'administrative data,' 'secondary uses of data,' and 'social science research.' It outlined recent developments that justify this thesis' scope. The analysis within this thesis suggests there are many powerful actors across a range of domains, including the political, commercial, and academic. These all seek to capitalise on the advances fostered by developments in data processing and its associated technologies. They may not share the same understanding of the potential risks involved or sufficiently consider the citizens' expectations. The small amount of empirical work in this area may have facilitated an understanding of individuals' attitudes and expectations rooted instead in these entities' needs. Such an approach risks undermining the potential benefits of reusing administrative data in research by playing to individuals' fears about modern data processing practices; the lack of control, the fear of being constantly

[109] Denis Galligan, *Law in Modern Society* (Oxford University Press 2010).

[110] Nonet and Selznick (n 92).

tracked, and pervasive mistrust in government and other institutions. This has implications for the lawfulness of these practices, where the law requires data users to demonstrate a certain level of transparency. There is a risk that by failing to engage with citizens, institutions may lack the transparency required by law and may lead individuals to become suspicious of their motives for reusing data.

1.3.1 Part I

Part I encompasses the legal analysis of the GDPR and DEA. The GDPR is relevant as it currently retains its legal force in the UK, despite the UK's withdrawal from the EU. The UK's legislation in this area, most notably the Data Protection Act 2018, [111] was specifically enacted to comply with many EU obligations. The analysis also includes the DEA's provisions that govern 'sharing for research purposes.'

The original intention of this thesis was to use the analysis of the GDPR alone. However, the restrictions triggered by the pandemic coincided with the planned start of the empirical work. The DEA provides further insights into legal approaches towards the permissible reuses of administrative data in social science research. In contrast to the GDPR, the DEA 2017 was not part of a commitment to strengthen individuals' rights in the face of the new challenges posed by advances in data processing practices but rather drew together a series of commitments made by the Conservative party during the 2015 UK general election. Part 5, which concerns the reuse of data in research, was introduced as part of a governmental push to make 'better use of data within government while

[111] Data Protection Act 2018.

taking security into full account.'[112] The inclusion of two different pieces of legislation also allows for a consideration of the legislative responses at a national and supra-national level.

A close analysis of these two legislative instruments responds to the first research question. The analytical approach draws upon theories that highlight the socially constructed nature of knowledge within a given context.[113] Doing so draws attention to the practices, actors, and discourses that shape not only the legislative response but also the 'problem' which they are intended to 'solve.' This analytical approach engenders an understanding of the constructed nature of the problem, facilitating a wider framing of the problem. This better accounts for broader social, political, and ethical issues that may be of influence. It also recognises and analyses the multiple forms of knowledge that were present during the legislative process.

Part I demonstrates the role of laws and lawmaking in constructing regulatory meanings, including conceptions of regulatory problems and solutions, and spaces, including regulatory actors. Practices during the legislative processes demonstrate attempts to imbue the law with a certain democratic and societal legitimacy through elected representatives and public consultations. Nevertheless, the subsequent discourses, perpetuated by more powerful actors during the legislative processes' later practices, undermined citizens' inputs. In doing so, Part I raises questions about the legitimacy of the legal governance of

[112] See the Explanatory Notes to the Digital Economy Act 2017, 12.

[113] Carol Bacchi and Susan Goodwin, *Poststructural Policy Analysis a Guide to Practice* (Palgrave Macmillan 2016).

the reuse of administrative law for research purposes. Calls for further control in the consultations before the GDPR appear to be reflected in the provisions that strengthen individuals' rights. However, a closer analysis of the provisions shaped by subsequent practices and discourses suggests the prominence of a market-making approach towards data protection law. The approach promoted within the Council of the EU, the discourses that promoted the idea of data processing advances as an opportunity, and the construction of data subjects as astute, informed individuals served to promote a permissive approach towards the reuse of data. The GDPR may also claim substantive legitimacy by constructing individuals as regulatory actors equipped with rights who are thus empowered by this reform. However, the analysis suggests this empowerment is illusory.

Similar pre-legislative practices happened before the DEA, yet the discourses focused on current 'barriers' to data reuse portrayed the UK as falling behind other countries. These discourses promote a permissive approach towards the reuse of data for certain kinds of research by emphasising the potential benefits of the outcomes of data processing. This focus overshadows the impact on individuals' interests. A focus on tangible benefits and criteria of suitability, premised upon statistics-based research, potentially carves out a subsection of research deemed worthy of greater access to administrative data.

Therefore, the analysis within Part I suggests that, despite public consultations and references to norms and societal expectations, subsequent discourses may have created legislative outcomes that fail to align with societal expectations. This leads to Part II of the thesis, which explores societal expectations further.

1.3.2 Part II

Part II commences with a 'bridge' that links this thesis' legal analysis and empirical work. The first chapter explores the theories that posit societal expectations matter in determining the interpretation of laws, which provides space for the law to adapt to shifts in society in response to technological developments. This includes Poort's[114] and Nonet and Selznick's[115] writings on the need for interaction between the legislature and society. These arguments justify this thesis' empirical work and open up options for responses to identified divergences beyond legal reform. It allows for a reflection upon the notion of the law's legitimacy in a context where the values, norms, and beliefs that shape peoples' views towards the appropriateness of data processing practices may be diverse and/or nascent.

The first chapter of Part II also introduces and justifies the methodological approach, which combines a survey and a series of focus groups in a sequential mixed-methods study. A sequential mixed-methods study consists of a research design whereby 'researchers implement the quantitative and qualitative strands in sequence to use follow-up qualitative data to elaborate, explain, or confirm initial quantitative results.'[116] An outline of the strengths and weaknesses of quantitative and qualitative methods helps explain why a mixed-methods study was preferred. The chapter also justifies the sequential approach of the empirical work that comprises an online survey followed by a series of online focus groups.

---

[114] Poort (n 96).

[115] Nonet and Selznick (n 92).

[116] Vicki Clark and Nataliya Ivankova, *Mixed Methods Research: A Guide to the Field* (Sage 2016), 106.

This section details the changes to the planned empirical work considering difficulties in undertaking in-person empirical research due to restrictions triggered by the global pandemic.

The empirical research, consisting of an online survey followed by a series of five online focus groups attended by 36 participants, responded to the second research question. The quantitative aspect of the empirical work enables the collection of many responses and suggests areas where participants' views differed from the legal approach. Further qualitative work helps to attain a greater depth of understanding. Online focus groups provide the space for further elaboration of the points raised by the survey through deliberation and reflection of their personal experiences. Annex A provides a detailed account of the methodology, Annex B provides a copy of the online survey, and Annex C provides the questioning route for the focus groups.

The following chapter provides the analysis of the quantitative research. The online survey took place between October and December 2020, and 302 UK-based participants completed it. The survey comprised of thirteen questions, including two free-text answers and a series of questions using a 5-point Likert scale to gauge respondents' views. The survey was disseminated through several channels, including callforparticipants.com, a website that advertises accredited university research needing respondents, my networks, and through contacting online groups. The survey sought insights into respondents' views towards different aspects of the reuse of administrative data in research. The questions were informed by analysing the legal provisions that highlighted several areas deemed important during the legislative process, and existing empirical research.

The findings suggest a lack of consensus in some areas as to what respondents found to be acceptable. This includes the types of administrative data that respondents would find acceptable to be reused in research, the identifiability of administrative data that could be acceptably shared with researchers, and the acceptability of different actors reusing the data for research purposes. The findings also identify potential differences between the legal approach and respondents' expectations. 70.6% of respondents found it unacceptable for identifiable data to be reused for research purposes. In contrast to the law's approach, respondents made distinctions based on the type of actor seeking to reuse the data.

Chapters seven and eight present the qualitative findings from a series of online focus groups that happened during 2021. The qualitative research aimed to provide a richer understanding of the views collected during the quantitative aspect of the thesis and an insight into how participants justify their views. In doing so, it shed light on divergences between the legal approach and societal expectations and investigate why these divergences may occur in the context of reusing administrative data in research. Two 'provocations' were created, based upon the survey findings and existing empirical research. These allowed participants to consider and discuss the different aspects of this data processing practice, such as the actors, purpose, identifiability of the information in a more contextually-sensitive way than possible through the survey. A series of open-ended questions, which gave the participants further freedom to explore the factors identified as relevant in influencing their perceptions of the appropriateness of reusing data in research, complemented the provocations.

The analysis of the empirical data drew upon Bacchi's framework to shed light on the prominent discourses within focus groups that shaped participants' problematisations. This approach enabled the identification of divergences from the problematisations identified in Part I. The findings suggest the presence of divergences, including the prominence of a risk-oriented discourse within the focus groups, assumptions about the low trust participants had in key actors, research as an activity in the public interest, and the influence of high-profile scandals in shaping several participants' negative perceptions. The prominent discourses in the focus groups generated different problematisations that led to different effects to those identified in Part I. Consequently, participants constructed a less permissive approach towards the reuse in data in research as appropriate. This more restrictive 'solution' as a response to participants' problematisations is present in participants' desires for greater individual control, more anonymity, and the criticism of law as insufficient to render an activity acceptable to participants.

In contrast to the analysis in Part I, the diversity in views meant it was only possible to identify prominent, rather than dominant, discourses that generated a range of effects. Part II reflects upon the implications of this finding. It may be difficult to speak of the law's legitimacy in this context as the lack of a single system of socially constructed norms, beliefs, values, and definitions renders it difficult to state what actions will be assumed to be appropriate. These findings are used in Part III to consider the potential responses to the identified divergences between the law analysed in Part I and the views analysed in Part II.

1.3.3 Part III

The impact of clashing discourses, leading to different views on what is appropriate in this context, is examined in Part III. Part III explores how the law can respond to such diversity. The findings from Part I and Part II are used as a springboard for exploring the third research question; 'what could be done when the law and citizens expectations are not always the same, in a context where the latter is supposed to justify the approach of the former?' This final research question speaks to deeper questions regarding the relationship between law and citizens in a society in the context of innovative and rapid technological developments. The altering of the law to better align with current expectations is not perceived as the most prudent way forward. This argument draws upon literature that argues the law should not always be reflective of societal norms. This section outlines the difficulties involved in legislating emerging technologies[117] and justifications for the law to act as a 'bulwark.'[118] Doing so reflects upon the law's ability to demonstrate its legitimacy where no single set of norms, beliefs, values, and definitions enable judgements regarding the appropriateness and propriety of actions to be made.

Part III draws upon the insights gained from the notion of 'responsive law'[119] and a 'social licence' to reject the idea that the law's enactment is the 'end point' by which one may judge its legitimacy. Laws need not necessarily create

[117] Brent Daniel Mittelstadt and Luciano Floridi, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2015) 22 Science and Engineering Ethics 303.

[118] Mark A. Edwards, 'The Alignment of Law and Norms: Of Mirrors, Bulwarks, and Pressure Valves' (2014) 10 FIU Law Review 19.

[119] Nonet and Selznick (n 92).

static institutional arrangements but may foster fluid regulatory spaces in which a range of actors can participate in a plethora of ways. Though the notions of 'expectations' and the 'public interest' currently risk interpretation through the lens of the needs and wants of data users, it could operate as a usefully flexible vehicle with which to accommodate the evolution of individual's views over time. Where the 'public interest' may be considered an overarching purpose of the law, the notion of responsive law enables this purpose to be 'held open to reassessment in the light of its consequences for the values at stake.'[120] The concept of a social licence can operate where there are meaningful differences between what is legally permissible and societally acceptable, helping to sustain and legitimise particular practices.[121] The creation and maintenance of a social licence require engagement with diverse perspectives and interests to help ensure that 'current and future practices are aligned with the values of society.'[122]

Therefore, the better inclusion of citizens in the decision-making processes that determine the permissibility of data processing activities is preferred to enable the pursuit of legitimacy. There are opportunities in the regulatory space created by the GDPR and DEA for the ex-post involvement of citizens. This inclusion could only be possible through individuals' ongoing and inclusive participation in determining the acceptability of the reuse of administrative data for social research. This participation should not be reduced to tick box exercises

[120] Ibid, 82.
[121] Mhairi Aitken and others, 'Establishing a Social Licence for Financial Technology: Reflections on the Role of the Private Sector in Pursuing Ethical Data Practices' (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720908892>.

[122] Ibid.

nor treat citizens as 'passive recipients of scientific knowledge.'[123] Instead, 'technical approaches need to be combined with social responses that build relationships of trust'[124] to respond to the challenges identified in this thesis.

The final chapter introduces some of the novel governance mechanisms that could facilitate this greater inclusion. It also explores whether this could increase the formal and substantive legitimacy of the law in this area as a regulatory mechanism. The chapter outlines four mechanisms, including bolstering citizens' views through their inclusion in approvals panels, using citizen's assemblies[125] to involve individuals further and forge links between them and data controllers, an even broader Taiwanese participatory practice,[126] and public data trusts.[127] This chapter presents the key findings and the contributions of this thesis to the field of socio-legal research. The limitations of the research and the potential direction of future research are also considered.

[123] Ibid.

[124] Ibid.

[125] Participedia, 'Citizens' Assembly' (2020) <https://participedia.net/method/4258>.

[126] Participedia, 'VTaiwan' (2021) <https://participedia.net/method/7387>.

[127] Marina Micheli and others, 'Emerging Models of Data Governance in the Age of Datafication' (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720948087>.

Part I: The Law

2 EXPLORING A LEGAL CONSTRUCTION OF THE PERMISSIBILITY OF THE REUSE OF DATA IN RESEARCH: THE GENERAL DATA PROTECTION REGULATION

2.1 INTRODUCTION

This chapter explores the General Data Protection Regulation (GDPR)[1] as a legislative response to the innovations in data processing outlined in the previous chapter. This draws upon Bacchi's[2] work and her framework for analysing the constructivist nature of policies, highlighting the socially constructed and context-dependent nature of knowledge.[3] The context of creating legislation may influence the understandings and constructions of 'problems.' Her argument that policy is 'inherently political' and that surrounding discourses create competing representations of political issues[4] are equally applicable to the legislative process.

As a mechanism for governance, the law uses the same understanding of power present in Bacchi's work. Law can operate as a 'productive force'[5] that shapes conceptions, perceptions, and preferences in such a way as to make certain

[1] EU General Data Protection Regulation (GDPR): 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.

[2] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

[3] Carol Bacchi and Susan Goodwin, *Poststructural Policy Analysis: A Guide to Practice* (Palgrave Macmillan, 2016).

[4] Carol Bacchi, *Women, Policy and Politics: The Construction of Policy Problems* (Sage 2011).

[5] Ibid.

ideas palatable and acceptable. Legislating entails similar dynamic processes of meaning-making as present in policymaking. The production and use of knowledge are 'interwoven with the generation of ideas about what the problem is and how it might be addressed.'[6]

Bacchi's starting point helps to identify the role and relevance of discursive influences by elucidating how the 'problem' of modern data processing was understood, constructed, and 'solved' by those involved in the legislative process. This approach disrupts certain aspects of the law that has a taken-for-granted status, investigating how certain practices shaped the treatment of data processing as an activity requiring legal regulation.[7] This analysis aligns with Foucauldian ideas that emphasise how our perceptions of a problem impact our construction of a solution,[8] a process defined as 'problematisation.'

This approach highlights how laws and regulations are not responses to a problem waiting to be discovered and solved, outside the legislative process, and nor are laws mere 'containers for words.'[9] The concept of 'problematisations' emphasises the constructed nature of problems and actors' roles and practices in shaping them. The actors include lawmakers and others of influence, including lobbyists, experts who contribute their knowledge, and the public. The practices

[6] Anna Wesselink and others, 'Technical Knowledge, Discursive Spaces and Politics at the Science–Policy Interface' (2013) 30 Environmental Science & Policy 1.

[7] Carol Bacchi, 'Why Study Problematizations? Making Politics Visible' (2012) 2 Open Journal of Political Science 1.

[8] Ibid.

[9] Lindsay Prior, 'Repositioning Documents in Social Research' (2008) 42 Sociology 821.

include 'what is said and done as part of the legislative process, where rules are imposed, where reasons are given.'[10] Lawmakers' decisions mean legislative texts can frame key terms in specific ways that are then disseminated and used within society. This chimes with Cotterrell's [11] views that the language present in legislative texts can shape individuals' thoughts towards specific terms and ideas. This better accounts for broader social, political, and ethical issues that may influence by recognising and analysing the multiple forms of knowledge present during the legislative process.

This chapter explores the GDPR's content and its impact on society to consider how it drives forward certain activities, such as reusing data in research. The following chapter complements this analysis by exploring the potential for the GDPR to generate certain effects within society, specifically the shaping of a certain attitude towards reusing administrative data in social science research. Faulkner notes that the discursive linguistic construction of new technologies and scientific domains is 'an important motor force in their emergence and steerage.'[12] This approach is reflected in acknowledgement that legislative texts have 'inscribed in them the outcome of socio-political and technoscientific debates and negotiations.'[13] This potentially sheds light on the effects of the legislative process. How did modern data processing become a problem that could only be solved by

[10] Bacchi (n 7).

[11] Roger Cotterrell, 'Why Must Legal Ideas Be Interpreted Sociologically?' (1998) 25 Journal of Law and Society 171.

[12] Alex Faulkner, 'Law's Performativities: Shaping the Emergence of Regenerative Medicine through European Union Legislation' (2012) 42 Social Studies of Science 753.
[13] Ibid.

the reform present in the GDPR? What was rendered permissible through the problematisation that justified this specific reform? What are its implications?

The analysis thus aligns with the point raised by Raab and Koops that

> *designing the instruments that are the 'how' of data protection is not a dispassionate technocratic process of choosing tools to do a job but a political process in which there are many conflicts and interests, in which more than data protection is at stake.*[14]

This statement raises an interesting point relating to the understanding of the law's legitimacy, as introduced in the previous chapter. This understanding links the law's legitimacy to its perception as an entity that acts in a desirable, proper, or appropriate manner within a system of socially constructed norms, values, beliefs, and definitions.[15] The reference to the political process also resonates with Solum's point that the exercise of such political power is justifiable where exercised in accordance with what 'citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational.'[16] Therefore, the interests that are amplified or only exist within the legislative process may undermine the legislative processes' claims to be acting desirably and appropriately.

[14] Charles Raab and Bart-Jan Koops, 'Privacy Actors, Performances, and the Future of Privacy Protection' in Serge Gutwirth, Yves Poullet and Paul Hert (eds), *Reinventing Data Protection?* (Springer 2009).

[15] Mark Suchman, 'Managing Legitimacy: Strategic and Institutional Approaches' (1995) 20 Academy of Management 571.

[16] Lawrence Solum, 'Public Legal Reason' (2006) 92 Virginia Law Review 1472.

2.1.1 Chapter Outline

The chapter first explores the initial justifications for the reforms that culminated in the GDPR to analyse how the problem representation of modern data processing arose. The first section focuses upon two aspects of these justifications. Firstly, the competing constructions of the role of rights in governing permissible uses of data and, secondly, the role of innovation in constructing the economic and societal interests engaged by novel developments. The first aspect of this justification explores the role and relevance of public consultations in shaping the problem representation. How far did these mechanisms incorporate societal views to secure societal legitimacy in light of a range of representations from various actors? The analysis interrogates how certain representations arose, clashed, and subsequently shaped solutions.[17]

Next, the chapter outlines developments that occurred during the drafting of the GDPR, exploring how the actors and practices influenced a certain problematisation that resulted in the GDPR as the solution. The developments relevant to reusing administrative data in social science research are the focus of this section. The following chapter then explores the effects of the GDPR as the legislative outcome for this data processing activity. Aligning with Bacchi's approach,[18] the analysis acknowledges that what is left 'unproblematic' by a problem representation is still essential to understanding the resulting solution. Doing so highlights the existence of 'silences' in the discourses that shaped the GDPR and alternate ways of thinking about the problem. More dominant

[17] Bacchi (n 7).

[18] Bacchi (n 3).

discourses marginalise competing discourses that may become 'silences' in this context. This chapter explores two 'silences' in the GDPR. These silences concern the compatibility between big data analytics and the GDPR's provisions and the appropriateness of individual rights as the best mechanism to protect the interests affected by data processing.

The conclusion explores the consequences of the identified problematisation. The analysis suggests the problematisation results in a presentation of an irresolvable tension between individual rights and a construction of the 'public interest' underpinned by economic interests. This results in uncertainty regarding the scope, nature, and operation of the new provisions. In the context of reusing administrative data in research, this justifies a broad range of activities as 'worthy' of a facilitative approach due to the activity's inherent benefits. The following chapter then explores the effects of the GDPR in further detail

2.2 Generating the Need for the General Data Protection Regulation

This chapter explores how legislative reform facilitated certain data processing practices, focussing upon reusing administrative data in social science research. This draws upon literature that conceptualises the relationship between legislative texts and innovations in constructivist terms through a sociolegal lens.[19] Therefore, it does not begin with an analysis of the legislative text but rather the 'problem' that triggered such a response as the 'solution.' Consequently,

[19] Bacchi (n 3).

the starting point must be the construction of the 'problem' of modern data processing to understand the GDPR as the 'solution' to regulating data processing. Only then can we understand how the GDPR sought to 'solve' the problem identified. By identifying the influence of different actors, processes, and narratives, this chapter interrogates how representations arose, clashed, and subsequently shaped solutions.

The texts shaping the GDPR serves as an 'entry point' to the discourses that shaped the problem that triggered the GDPR as the legislative solution. This analysis identifies different and sometimes competing discourses that frame the 'best' legislative response. The term 'framing' denotes discourses that organise knowledge and experience that both 'constitute and contribute to the process of conveying ideas and values.'[20] This chapter explores the implications of a market-based framing, which promotes economic interests, and a human-rights-based framing, which promotes the protection of fundamental rights. How did these framings influence a facilitative approach towards reusing administrative data in social science research? On the surface, references to 'reasonable expectations,' the 'public interest,' and the prominence accorded to the rights-based protections suggest the dominance of a human rights-based framing. A human-rights based framing is not concerned with market rationalities and the economic aims of the EU. Instead, it serves to secure the social aims of the EU and protect individuals' rights as expressed in the sources of rights within the EU. However, the findings presented in this chapter cast doubt on this claim. Instead, the problematisation

[20] Gordon Bache, Mark Flear and Tamara Hervey, 'The Defining Features of the European Union's Approach to Regulating New Health Technologies" in Mark Flear and others (eds), *European Law and New Health Technologies* (Oxford University Press, 2013).

and the discourses surrounding the legislative process placed data protection within a neoliberal frame, prioritising the economic aims of the EU. A neoliberal framing operates to disseminate market rationalities beyond market spheres, meaning 'good' social policies align with economic optimisation.[21]

The findings have implications for the legal approach to reuses of administrative data in social science research. The underlying justifications for the legislative approach, including references to 'reasonable expectations' and the 'public interest', are infused with a neoliberal characterisation. This fosters economic interests rather than human rights or other matters that may be characterised as public interests. But how does this influence the rights-based protections present in the GDPR? The following section identifies and analyses the different constructions of the role of rights in the texts published during the initial proposal stage of the GDPR.

2.2.1 Competing Constructions of the Role of Rights

As the only institution empowered to initiate legislation,[22] the Commission issued a Communication in 2012 entitled 'Safeguarding Privacy in a Connected World: A European Data Protection Framework for the 21st Century.'[23] This document officially triggered the legislative process. Nevertheless, a fuller understanding of the problematisation at hand, its underlying assumptions, and how it arose

[21] Mark Flear, 'The Open Method of Coordination on Health Care after the Lisbon Strategy II: Towards a Neoliberal Framing' (2009) 13 European Integration Online Papers 1.

[22] See Article 17, Consolidated version of the Treaty on European Union OJ C326/1.

[23] Commission, 'Safeguarding Privacy in a Connected World a European Data Protection Framework for the 21st Century' (Communication) COM (2012) 9 final, 25 January 2012.

requires the analysis of preceding events. The consultations before the Communication are therefore relevant to understanding the GDPR's approach. The consultations aimed to inform the approach of the Commission's proposals to reform data protection law. The intention was to incorporate the input of the general public, Member State authorities, private stakeholders, and national data protection authorities whose views were collected during the consultation process. This practice sheds light on the problematisation triggering the GDPR 'solution' and some of its underlying assumptions. In seeking responses from a broad range of interested parties, the consultation process emphasised the range of interests potentially affected by the current law and subsequent reforms. The problematisation thus requires the input of more than the 'traditional' actors involved in the legislative process. This practice suggests there is a role for the general public to play. This constructs the legislative process as an inclusive practice that considers diverse views and genuinely attempts to find a consensus by better reflecting the norms, beliefs, values, and definitions used by those within society to determine the appropriateness of something. The underlying assumption is that this inclusivity helps to legitimise any legislative outcome. Therefore, in shaping how the GDPR will be regarded, the consultation efforts attempt to imbue the law with a heightened societal legitimacy. This justifies its approach to data processing as reflecting the interests of all those affected.

The consultation demonstrated the existence of competing representations concerning the 'problem' of current data processing practices to which legal reform could be a solution. Individuals' responses focused upon the impact upon individuals' rights, attributing issues to their perceived loss of control over their

data. These representations thus presume that individual rights in the context of data protection enable an appropriate control level. This may explain why many responses focussed on the importance of consent and anonymity as prerequisites for permissible data processing. Individuals often grounded their views in their perceptions of current threats to their interests, including government monitoring and big data analytics, to justify their need for more control. One stated that, currently, '[t]here is a mismatch of objectives between the rights of the individual and the rights of the state and superstate' and, under the current Directive, they perceived an 'intention to weaken the rights of the individual in favour of interstate utility.'[24] Another submission problematised the developments in terms of 'challenges for personal control in that individuals have very little power over how their personal information is used, especially once it has been passed to a third party.'[25]

In contrast, organisations used the consultations to emphasise that 'data protection must always be balanced against other fundamental rights such as the right to property and freedom of economic activity'[26] i.e., the rights that support commercial and economic interests. These comments indicate that the beneficial nature of rights-based protections was constructed to encourage trust in data

[24] Nick Fluck, 'Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data - Citizen' (2009) <https://eceuropaeu/home-affairs/sites/homeaffairs/files/what-is-new/public-consultation/2009/pdf/contributions/citizens/fluck_nick_enpdf>.

[25] Nadja Kanellopoulou, 'European Commission Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data EnCoRe Project Response' (2009).

[26] Directorate-General Justice of the European Commission, 'Summary of Replies to the Public Consultation about the Future Legal Framework for Protecting Personal Data' (2010) <https://web.archive.org/web/20111117000630/http://ec.europa.eu/justice/news/consulting_public/0003/summary_replies_en.pdf>, 3.

processing practices. This construction links rights with individuals being more willing and able participants in the internal market and is suggestive of a market-based framing of the problem. DIGITALEUROPE, a coalition of commercial entities, viewed the aim of data protection law as the 'creation of an environment where users can further trust their use of technology, fostering technological innovation and enabling continued consumer benefits.'[27] Minkkinen[28] notes that the rhetoric of balancing competitiveness and data protection, which is central to this discourse, promotes internal market objectives. Public authorities adopted a similar approach, stating that 'existing standards might be inadequate for dealing with new technologies.'[29] Whilst the importance of rights-based protection in this context was not undesired, 'the right to data protection will sometimes need to be balanced with other equally important fundamental rights.'[30] The UK's Ministry of Justice highlighted the need to ensure any reform 'remains appropriate alongside our increased processing ability…[and] also ensure that changes in legislation support the public benefits that can be delivered by using personal data alongside changing technology.'[31]

[27] DIGITALEUROPE, 'Response to European Commission Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data' (2009).

[28] Matti Minkkinen, 'Making the Future by Using the Future: A Study on Influencing Privacy Protection Rules through Anticipatory Storylines' (2018) 21 New Media & Society 984.

[29] Directorate-General Justice of the European Commission (n 26), 12.

[30] Ibid.

[31] Ministry of Justice, 'UK Response to the European Commission Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data' (2010).

These competing representations influenced the legislative solution, including the approach towards reusing data in research. The consultation responses demonstrate the importance and popularity of 'rights' across the spectrum of potentially impacted parties. Their subsequent inclusion thus serves to legitimise a legislative solution. Nevertheless, one must carefully examine to what end rights are used and how they support certain discourses. The findings demonstrate that the competing representations of 'rights' and their different underlying purposes may appear to facilitate a human rights-focussed approach to reusing data in research. However, the closer examination undertaken in this thesis reveals the prominent presence of market-based discourses. In the context of consultations, this could provide societal legitimacy for a legislative response that fails to align with individuals' expectations towards the role of rights in data processing. Therefore, the reusing of data in social science research may align with societal expectations. However, this section suggests the presence of market-based discourses as shaping the protections offered by rights. Doing so potentially undermines the law's legitimacy in governing reuses of data by promoting norms, beliefs, and values that citizens do not use in their determination of what is appropriate in this context.[32] This may cause divergences between what becomes lawful and societally acceptable.

[32] Suchman (n 15).

2.2.2 The Role of Innovation in Constructing Economic and Societal Interests.

The Commission Communication repeatedly emphasised the impact of data processing by reference to its novelty and capacity to enable innovation. It perceived the 'rapid pace of technological change' as having 'profoundly changed' the landscape within which personal data are collected, accessed, and used.[33] The framing of technology, emphasising novelty and advancements, played an essential role in how those responsible for drafting the GDPR responded. [34] Perceptions of technology are significant where, in the political arena, 'attitudes towards a technology can matter more than any of its intrinsic qualities.'[35] This supports the argument that 'problems' do not 'sit outside the policy process, waiting to be addressed and fixed.' [36] The approach within the Commission Communication resonates with what Frahm and others identify as a 'deficit model of innovation' that 'presents societies and institutions as lacking innovativeness and novel products such that interventions framed as innovation can be offered as the solution.'[37]

---

[33] Commission (n 23).

[34] Elen Stokes, 'Something Old, Something New, Something Borrowed: Emerging Health Technologies and the Continuing Role of Existing Regulations' in Mark L Flear and others (eds), *European Law and New Health Technologies* (Oxford University Press 2013).

[35] Ibid.

[36] Bacchi (n 3).

[37] Nina Frahm, Tess Doezema and Sebastian Pfotenhauer, 'Fixing Technology with Society: The Coproduction of Democratic Deficits and Responsible Innovation at the OECD and the European Commission' [2021] Science, Technology, & Human Values <https://journals.sagepub.com/doi/full/10.1177/0162243921999100>.

This focus on innovation problematises the existing law as unable to determine what should and should not be lawful. This is evident in the Data Protection Directive being problematised as failing to provide

> *neither the degree of harmonisation required, nor the necessary efficiency to ensure the right to personal data protection in today's new, challenging digital environment.*[38]

By highlighting the negative impact on both economic and societal interests, the Commission constructs a wide-ranging problem faced by the EU and Member States. However, it does not overtly consider whether these interests will necessarily align towards a single solution or if they might pull in different directions through contradictory responses. Furthermore, the emphasis on societal interests, while not acknowledging their potential tension with economic interests, constructs the GDPR's approach as promoting societal interests. Nevertheless, tensions between societal and economic interests served by data processing may lead to a legislative 'solution' that does not promote societal interests in a way expected by individuals. Instead, this tension becomes a 'silence' that remains unexplored by the dominant discourses during the legislative process. Furthermore, the Communication resonates with the works of Fougère and others[39] . They highlight how certain framings can present 'solutions' as pragmatic and well-adapted, regardless of ideology as a means to conceal antagonism. Doing

[38] Commission (n 23).

[39] Martin Fougère, Beata Segercrantz and Hannele Seeck, 'A Critical Reading of the European Union's Social Innovation Policy Discourse: (Re)Legitimizing Neoliberalism' (2017) 24 Organization 819.

so means that this framing excludes alternative ways of addressing the identified challenges and prevents debate over the potential causes of the challenges.

The language of innovation is also present in discourses that construct legal reform as an opportunity for the EU to be a world leader in this field. The Communication portrays legislative reform as a chance to ensure 'effective and consistent protection in a…globalised environment.'[40] This constructs the EU as in a prime position to influence the approach of other institutions and nations. The European Data Protection Supervisor's views on what should be permissible, based upon potential competition from parallel developments within other international organisations, including the Council of Europe and the OECD, shapes the problematisation.[41] The emphasis on the need for a prompt response, which considers the global implications of data processing, increases the opportunity for certain EU principles to be part of a possible 'binding global instrument.'[42] Legislative reform could help secure the EU's goal of being 'the most dynamic and competitive, knowledge-based economy in the world.' [43] The comments made by influential actors demonstrate that the drive for globalisation

[40] Commission (n 23).

[41] European Data Protection Supervisor, 'Opinion of the European Data Protection Supervisor on the Data Protection Reform Package' (2012) <https://edps.europa.eu/sites/default/files/publication/12-03-07_edps_reform_package_en.pdf>.

[42] Ibid, 6.

[43] European Council, 'Conclusions of the Presidency' (2000) <https://www.europarl.europa.eu/summits/lis1_en.htm#:~:text=PRESIDENCY%20CONCLUSIONS&text=The%20European%20Council%20held%20a,of%20a%20knowledge%2Dbased%20economy.>.

is framed in terms of its market-making possibilities rather than a human-rights-based framing. The EU Digital Agenda Commissioner stated that

> *[r]ules must take account of the impact on businesses as well as on citizens…faced with too many restrictive rules and obligations, would-be data controllers may just take their bright ideas outside the EU.*[44]

This drives forward a legislative solution that may diverge from societal expectations towards reusing their data, depending on how far they perceive such aims as a legitimate aim of reforming data protection laws.

A market-making framing is not the most prominent discourse present in particular views that facilitated a permissive governance approach to reusing data. The Chair of the EU's Working Party on Statistics, one of the preparatory bodies of the Council of the EU, highlighted that the proposed reforms could increase burdens on research institutions. This justification focused on the risk of a 'serious deterioration of the quality and coverage of statistics and research'[45] instead of the economic consequences. Nevertheless, it facilitated the market-based discourses and free flow of data rather than the rights-based discourses. This discourse constructs the rules relating to the rights to be informed and access to personal data, and the use of delegated acts to specify the requirements for data processing for research purposes, as 'disastrous' administrative burdens.[46] The

[44] Neelie Kroes, 'Cloud Computing and Data Protection Reform' (2012) <https://ec.europa.eu/archives/commission_2010-2014/kroes/en/blog/cloud-data-protection.html>.

[45] Chair of the Council Working Party on Statistics to The Working Group on Information Exchange and Data Protection, 'Proposals for Changes in the General Data Protection Regulation' (2012).

[46] Ibid.

introduction of this discourse at the start of the legislative process, by an actor whose influence arises from their expertise in research rather than their commercial interests, potentially supports the discourses that promote a permissive approach to data processing.

It remains unclear whether the expectations of citizens and their views towards the public interest in research would justify fewer protections of their individual interests to facilitate research. Nevertheless, the emphasis on control and anonymity, present in citizens' consultation responses, appears to place these discourses at odds. The Commission's Communication presents a different view. It recognises both the opportunities presented by new possibilities and the 'unprecedented scale' of processing as increasing the 'risk to individuals' rights' that 'challenges their capacity for keeping control over their data.'[47] Doing so promoted a solution that sought to ensure the more market-based benefits were not unduly constrained. However, such a solution could be constructed in a plethora of ways, depending upon the preferred reconciliation of these interests. This section highlights how competing discourses can exist, and the next section will consider how the power wielded by certain actors and institutions promoted a certain discourse. This demonstrates how the power of certain actors may have shifted the legislative response in the GDPR further from societal expectations.

[47] Commission (n 23).

2.3 DEVELOPING THE GENERAL DATA PROTECTION REGULATION

This section explores several developments that occurred during the legislative process of the GDPR. It considers the role of lobbyists during the drafting process, external events that influenced the legislative outcome, and the impact of the institutions within the EU upon the construction of the drafts and final version of the GDPR.

2.3.1 The Role of Lobbyists

The impact of lobbying is relevant to understanding how the legislative process may have prompted a construction of permissible data processing practices that diverge from societal expectations. The concept of lobbying, whereby external parties seek to influence legal reform or other policies through contact with parties and actors involved in the legislative process, is not a new or unique practice arising during the GDPR. Nor is it correct to surmise that *any* lobbying is detrimental to the democratic process. Lobbying can play a legitimate role where those involved in the lawmaking process may not be cognisant of all the information required to make informed decisions or where there is a complex background to the proposals.[48]

[48] Doris Dialer and Margarethe Richter, 'Lobbying in Europe: Professionals, Politicians, and Institutions under General Suspicion?' in Doris Dialer and Margarethe Richter (eds), *Lobbying in the European Union* (Springer 2019).

However, where lobbying operates without 'democratic scrutiny, transparency, and ethical standards'[49] and reaches previously unknown levels,[50] there is scope for the dominant discourse to reflect lobbyists' desires rather than societal expectations. Moreover, as commercial entities were responsible for most of the lobbying, these actors likely promoted more market-framed constructions of data protection laws. The amendments proposed by Amazon, the American Chamber of Commerce to the European Union, and eBay, which matched the Committees of the European Parliament's amendments,[51] support this claim. The GDPR's development was of interest to American firms. In contrast to the EU's perceived focus on fundamental rights, the American approach to data protection suggests a preference for self-regulation. This approach prioritises individuals' and commercial entities' freedoms. The opposition of the EU Home Affairs Commissioner to the proposals provides a further indication of the pressures placed on certain actors within the legislative process. Some attributed her opposition to her attempts to enable smoother negotiations with the US on data transfers, which were also part of the planned legislative reforms.[52] Therefore, such opposition may have potentially influenced the problematisation and dominance of certain discourses.

[49] Ibid, 5.

[50] James Fontanella-Khan, 'Brussels: Astroturfing Takes Root' *Financial Times* (June 26, 2013) <https://www.ft.com/content/74271926-dd9f-11e2-a756-00144feab7de>.

[51] 'Europe-v-Facebook.org' (2013) <http://europe-v-facebook.org> accessed March 9, 2021.

[52] EUROACTIV.com, 'Commission Strife Risks Delaying Data Protection Overhaul' (2012) <https://www.euractiv.com/section/justice-home-affairs/news/commission-strife-risks-delaying-data-protection-overhaul>.

These influences provide insights into why certain aspects of overarching approaches of the GDPR may fail to align with societal expectations. By reinforcing the dominance of market-based discourses the influence of lobbyists may have engendered a legislative approach that diverges from societal expectations and understandings of what should be legally permissible. Lobbyists are not the only potential external influence upon the legislative drafting process. The following section considers the impact of an unforeseen event on the legislative process.

2.3.2 The Impact of External Global Events

Not all influences upon the legislative process are intentional acts by interested actors. Sometimes, external events introduce new justifications or reasons for action, undermine the attractiveness of a certain discourse, or promote a previously silenced discourse. During the legislative process, Edward Snowden, a contractor at the US National Security Agency, released many documents. These detailed the surveillance undertaken by the USA, including the tapping of several European Union buildings. The leak detailed the PRISM program, which allowed the USA's National Security Agency a 'backdoor' into data collected by firms including Facebook and Google.[53] Such revelations about the intrusive and secretive nature of governmental data collection, storage, and other processing activities significantly affected the discourses during the legislative process.

[53] BBC News, 'Edward Snowden: Leaks That Exposed US Spy Programme' *BBC News* (January 17, 2014) <https://www.bbc.co.uk/news/world-us-canada-23123964>.

The Snowden affair aligns with Birkland's definition of a 'focusing event,' which is 'sudden, relatively uncommon, can be reasonably defined as harmful (…) and that is known to policymakers and the public simultaneously.'[54] The scandal's notoriety and the resulting perceived failings of current laws and policies accelerated the drive for a 'solution.' This increased attention can 'further tilt the balance of debate in favour of pro-change groups.'[55] The pro-change groups here focussed on promoting the human rights-based discourses relevant to data processing. Research by Kalyanpur and Newman[56] suggests the effects on public discourse were 'immediate and long-lasting, shifting the ground of the GDPR debate.' The month of the scandal saw triple the number of articles concerning privacy compared to the previous month. This increase in privacy-related news stories was still evident six months later.[57] The Snowden revelations prompted several significant 'aftershocks'[58] that retained the GDPR's salience throughout the negotiations, including a series of high-profile court cases that rallied against perceived privacy intrusions.[59]

[54] Thomas A Birkland, 'Focusing Events, Mobilization, and Agenda Setting' (1998) 18 Journal of Public Policy 53.

[55] Ibid.

[56] Nikhil Kalyanpur and Abraham L Newman, 'The MNC-Coalition Paradox: Issue Salience, Foreign Firms and the General Data Protection Regulation' (2019) 57 JCMS: Journal of Common Market Studies 448.

[57] Ibid.

[58] Ibid.

[59] See, for example, Case C-362/14 Maximillian Schrems v Data Protection Commissioner [2015] ECR I-650.

This scandal promoted discourses that emphasised the role of 'the public interest' in determining what data processing should be legally permissible. The strength of public feeling and the prominence of the proposed legal reforms in the public eye led many politicians to be more overtly supportive of the rights-based framings of the proposed reforms. The German chancellor was initially reticent in supporting the GDPR, but the revelations coincided with her becoming more vocally supportive towards a rigorous legislative approach.[60] A second scandal, arising from certain Members of European Parliament (MEPs) copying and pasting amendments proposed by commercial entities, prompted a further public backlash. These events may explain the significant differences between the original draft of the GDPR and the European Parliament's first reading. Such changes align with Rossi's point that 'business power decreases as political salience goes up.'[61] When the public is relatively unaware or appears not to care about an issue, this creates space for an actor involved in the legislative process to follow the self-interested advice of corporations. This becomes increasingly untenable where public salience increases and where actors must look to additional sources of information. These discourses, rooted in bolstering rights protections and preventing intrusions perceived as excessive, align with a more rights-framed approach to the 'public interest.' The Snowden revelations demonstrate the impact of external events upon the GDPR's problematisation and

[60] Melissa Eddy and James Kanter, 'Merkel Urges Europe to Tighten Internet Safeguards' *The New York Times* (July 16, 2013) <https://www.nytimes.com/2013/07/16/world/europe/merkel-urges-europe-to-tighten-internet-safeguards.html>.

[61] Agustín Rossi, 'How the Snowden Revelations Saved the EU General Data Protection Regulation' (2018) 53 The International Spectator 95.

the proposals' continued malleability. The focus shifted to how the GDPR could respond to these new challenges to individuals' interests and rights and its potential for constraining the novel inappropriate data processing practices those innovations made possible.

The Snowden revelations did not directly concern the data processing activity of research *per se.* However, it justified the rights-based approach taken by the Committee on Civil Liberties, Justice and Home Affairs (LIBE Committee) in their draft proposal of the GDPR. A later section concerning the Committee system within the European Parliament details the Committee's impact on the GDPR's development. This section focuses specifically upon the impact of these external global events on the development of the GDPR. The rights-based approach promoted by the LIBE Committee had significant implications for secondary uses of data in research. Nearly 100 non-commercial research organisations signed a position paper to argue the proposed amendments 'would make much research involving personal data at worst illegal and at best unworkable.'[62] The paper highlighted the types of research that could be in jeopardy. This included research concerning common chronic diseases, such as Alzheimer's and cancer, and research that could 'inform policy in areas such as health inequalities and economic instability.'[63] The need to demonstrate a 'high public interest' was criticised for being 'problematic' as the 'results and impact of

[62] Wellcome Trust, 'Protecting Health and Scientific Research in the Data Protection Regulation (2012/0011(COD)) Position of Non-Commercial Research Organisations and Academics' (2014) <https://www.russellgroup.ac.uk/media/5288/dpr-ep-position-joint-statement-april-2014.pdf>.

[63] Ibid.

the study are not known at the outset.'[64] The requirement that Member States only permit exceptions to the requirement of data subject consent where 'that research cannot possibly be carried out otherwise' was condemned. It was constructed as creating a test that did

> *not take into account the nature of research [as] it is not clear how one could show that a particular research project 'cannot possibly be carried out otherwise' where it is theoretically possible, but not practicable (because of the sample size needed), to seek consent.*[65]

Moreover, a need to rely on Member State exceptions undermined the 'intended uniformity of research-related legislation throughout Europe…maintaining existing obstacles for free movement of research data across European borders.' [66] Within the Council of the European Union, Danish representatives warned the amendments would have 'catastrophic consequences for register-based research' and presented the proposals as a backwards step for data processing in research.[67] The criticisms were not just premised on the impact of the amendments on delivering public benefits. These non-commercial actors also raised economic interests,[68] including the risks posed to 'millions of euros of

[64] Ibid.

[65] Ibid.

[66] Olof Nyrén, Magnus Stenbeck and Henrik Grönberg, 'The European Parliament Proposal for the New EU General Data Protection Regulation May Severely Restrict European Epidemiological Research' (2014) 29 European Journal of Epidemiology 227.

[67] General Secretariat of the Council, 'Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - Comments on Articles 83a, 83b and 83c' (2013) <https://data.consilium.europa.eu/doc/document/ST-14210-2013-REV-4/en/pdf'>.

[68] Johannes Starkbaum and Ulrike Felt, 'Negotiating the Reuse of Health-Data: Research, Big Data, and the European General Data Protection Regulation' (2019) 6 Big Data & Society.

investment in scientific infrastructure, including cancer registries, cohort studies and biobanks.'[69]

This section demonstrates the differing understandings ascribed to the critical term of the 'public interest.' The amendments proposed by the LIBE Committee, justified by reference to the need to protect individuals' rights, did not promote research as an activity in the public interest that warranted a facilitative legislative response. Instead, echoing some of the initial consultation responses, the amendments focussed on individual control and the relevance of anonymity. When confronted with the implications of this approach for undertaking research, the definition of the 'public interest' and the appropriate response to innovative data processing practices retreated from the LIBE committee's approach. This 'retreat' raises questions as to how far the understanding of the 'public interest' that dominated the legislative process thus aligns with societal expectations. The shift away from personal control and towards an inclusive definition of research, incorporating research activities undertaken by commercial actors and research without a demonstrable benefit to the public, indicates a possible divergence in approaches.

The next section focuses on the institutions and actors involved in the drafting process of the GDPR rather than external events. It explores the EU's structure and institutions and certain actors' actions to investigate their role in promoting certain discourses and framings present in the GDPR.

[69] Wellcome Trust, 'Data Protection' *www.thetimes.co.uk* (January 2014) <https://www.thetimes.co.uk/article/data-protection-dj89jl8rb8w>.

2.3.3 The Impact of the Institutions Within the EU

The EU's co-decision system means that all the actors within the tripartite political system contribute to the final legislative document. The European Commission proposes the legislation, as demonstrated in the Communication analysed in this chapter. The other two EU institutions critical to the EU's legislative process are the European Parliament and the Council of the European Union. This structure means any legislation passed by this mechanism results from a significant collective effort, often spanning years. The European Parliament consists of 751 directly elected members. Under the ordinary legislative procedure, it cannot proper legislation but can approve, amend, or reject legislative proposals the Commission makes. The Council of the European Union represents the national governments of Member States, with its votes weighted according to the respective populations of the Member States. It also has the power to amend and accept legislative proposals made by the European Commission.[70] Through the proposed amendments and compromise positions, the interactions between the European Parliament and the Council of the European Union are important. They facilitate a richer understanding of the competing discourses that shaped the problematisation of modern data processing practices and the GDPR as the legislative 'solution.' The solution must be acknowledged as the outcome at a certain point in time and the result of intense debates and

[70] Paul Craig and Gráinne De Búrca, *EU Law: Text, Cases, and Materials* (Oxford University Press 2020).

conflicts. Certain phrases, key terms, and approaches were hotly contested and subject to many amendments during the legislative process.

2.3.3.1 Member State Influence in the Council of the EU

One feature of the Council of the European Union is the presidency, which is responsible for 'driving forward the Council's work on EU legislation, ensuring the continuity of the EU agenda, orderly legislative processes, and cooperation among the Member States.'[71] The presidency is a position held by a national government, rather than an individual, and rotates through the Member States on a biannual basis. The Member State holding the presidency has the power to influence the legislative process by operating as an 'honest broker'[72] between those within the Council and the different institutions of the EU.

The presidency's influence on the framing of the legislative process may partly explain the prominence of research, as a specific data processing practice, within the initial draft associated with the 2012 Danish presidency. Denmark is at the forefront of epidemiological research due to the existence of registries that include substantial amounts of information on all Danes[73] that can be accessed for research purposes. Such an approach towards research suggests that its potential benefits are strongly and widely recognised as in the public interest within

[71] European Council and Council of the European Union, 'The Presidency of the Council of the EU' (2019) <https://www.consilium.europa.eu/en/council-eu/presidency-council-eu>.

[72] Council of the EU and the European Council, 'The Presidency of the Council of the EU' (2021) <https://www.consilium.europa.eu/en/council-eu/presidency-council-eu/>.

[73] Study in Denmark, 'Denmark Launches World-Class Biobank' *studyindenmark.dk* (2019) <https://studyindenmark.dk/news/denmark-launches-world-class-biobank> accessed March 9, 2021.

Denmark. However, it does not mean that all Member States, and citizens residing within them, expect their data to be used in this way.

The developments during the 2013 Irish presidency further demonstrate the potential for the presidency to influence the legislative process. There is evidence of further support for the market framing discourses that promoted a more permissive approach to data processing during this presidency. Whilst acknowledging the need for new legislation 'which both protect citizens' rights and facilitate business,' the emphasis was on the 'enormous economic value' of data processing and how its insights have become 'essential for business and legislation.'[74] The need for rights-based protections, how these will be facilitated, and why they are essential is barely discussed. Furthermore, economic considerations are the sole justification for rights-based protections whereby

> *violations of data protection rules can have an enormous cost… Making sure customers are notified fast when their data is hacked will generate trust. More people will do business online: good for growth and good for jobs.*[75]

Based on this construction, there is a drive towards ensuring businesses within the EU can 'take advantage of this new computing and information-sharing landscape' with the overall goal of making the 'EU common market for goods and services more accessible to businesses.'[76] The acknowledgement of the impact on individuals nevertheless promotes a market-framing solution founded upon

[74] Viviane Reding and Alan Shatter, 'What EU Data Protection Can Do for Growth' (2013) <http://www.eu2013.ie/news/news-items/20130129shatterredingoped>.

[75] Ibid.

[76] Ibid.

improving individuals' trust and confidence in businesses. The purpose is to encourage their custom, supporting the economic goals of the EU. The country holding the presidency may help to explain the dominance of the market-based discourses that promote a permissible approach. Ireland is a poster child for 'light tough' regulations to encourage businesses to base their activities there.[77] Their presence ensures significant revenue for Ireland through the taxes they pay and the jobs they provide, a position Ireland would no doubt prefer to continue under the GDPR. The GDPR is a Regulation, rather than a Directive, so its provisions are directly applicable to the Member States, precluding variance at the Member State level. Therefore, only through business-oriented amendments of the entire GDPR could Ireland seek to maintain their privileged position. However, this does not bear any relevance to societal expectations towards the public interest in reusing data in research.

Alongside the presidency, the agreement of Member States within the Council was also influential in promoting certain framings during the drafting process. This input provides a further example of an apparent promotion of a rights-based framing of data processing that potentially belies its actual impact. One proposed amendment placed the Recital that acknowledged the specific existence of the right to the protection of personal data in a more prominent position within the proposal. This was due to the 'importance of the fundamental rights dimension of data protection in connection with other fundamental rights.'[78]

[77] Kalyanpur and Newman (n 56).

[78] The Presidency of the Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection

However, the Recital also reaffirms the non-absolute nature of the right. The right's applicability and force are limited by the need to consider the right's 'function in society,' a term that remains undefined, and the need for it to be 'balanced with other fundamental rights.' Though Recitals do not have independent legal value, they can help determine the nature of a provision and set out the reasons for the articles included within a legislative act.[79] The lens of the dominant market-based discourses suggests a construction of a 'function in society' that facilitates a more permissive approach towards data processing. The lack of detail about the actors involved in the processing means there is no delineation between research institutions and commercial entities. This suggests it is not relevant to the legal approach towards the permissibility of data processing, but empirical work does not support this approach.[80] This further exemplifies how market-framing discourses may lead to divergences between legal permissibility and societal expectations. The dominance of market-based discourses also explains other proposed amendments. These include a clarification that the right to the protection of personal data was not of a higher status than other rights, such as those that promoted economic freedoms and a specific reference to the EU's Internal Market as one of the legislation's aims. [81]

Regulation)' (2012) <https://www.statewatch.org/media/documents/news/2012/jun/eu-council-revised-dp-position-11326-12.pdf>.

[79] Europa Publications Office, 'Interinstitutional Style Guide: 2.2 Preamble (Citations and Recitals)' (2019) <https://publications.europa.eu/code/en/en-120200.htm>.

[80] Daniel Cameron, Sarah Pope and Michael Clemence, 'Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes' (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/>.

[81] The Presidency of the Council of the European Union (n 78).

The strong reluctance of Member States to impose 'excessive burdens' upon small and medium-sized enterprises is a further impact of their influence which promoted a market-based, permissive approach towards data processing. By representing the problem as the 'burden' placed on small business owners, the emphasis is upon what duties it is appropriate to place on them, rather than the rights and freedoms of individuals. Further supporting this view is the recognition that possible detriments faced by individuals must not only be considered in relation to the 'elements that are likely to influence the probability that a danger may actually materialise' but also the 'entrepreneurial freedom of the private sector and the tasks of general interest which public authorities carry out.'[82] The increasing dominance of the risk-based approach is also present in the tendency to require 'appropriate' measures to be in place. Such discourses foster a focus on the economic factors that would inhibit full participation in the Internal Market as the basis of decisions regarding the permissibility of data processing.[83]

During the 2013 Irish presidency, the risk-based approach dominated as the most appropriate approach towards reconciling the rights-based protection and market-based opportunities that the legislation sought to promote. The drafts permitted the moderation of the stringency of certain measures designed to protect individuals' rights based on the level of risk posed by a given data processing

[82] The Presidency of the Council of the European Union, 'Data Protection Package - Report on Progress Achieved under the Cyprus Presidency' (2012) <https://www.consilium.europa.eu/en/documents-publications/public-register/public-register-search/results/?AllLanguagesSearch=False&OnlyPublicDocuments=False&DocumentNumber=16525%2F12&DocumentLanguage=en>.

[83] The Presidency of the Council of the European Union, 'Implementation of Risk-Based Approach in the General Data Protection Regulation' (2013) <https://data.consilium.europa.eu/doc/document/ST-5702-2013-INIT/en/pdf>.

practice. This enables data processors to consider the nature, context, scope, and purposes of processing activities and the risks to the rights and freedoms of data subjects in their determination of the 'appropriate' measures to take. Doing so minimises the importance of the rights of individuals, making rights one of a range of factors to be considered.[84] In terms of 'reasonable expectations,' this challenges how far this approach tallies with citizens' expectations towards potential uses of their data. Do individuals expect the notion of 'risk' to influence the stringency of protecting their rights? As far as data reuse for research purposes is concerned, do citizens find that risks are acceptable when processing for this purpose? Can the notion of 'risk' as understood here be deemed compatible with citizens' understandings of the term 'public interest'?

The interactions between the market-framing approach and research are also evident in the legislative drafts proposed by the presidencies. The 2014 Italian presidency removed the need for research processing to be in the 'public interest.'[85] The presidency emphasised that this could exclude certain private actors that could nevertheless constitute 'important sources of our common cultural heritage.'[86] Whilst this stressed the need to support private institutions as a source of cultural importance, it simultaneously downplayed their commercial interests. This enables the institutions that pursue commercial interests to benefit

[84] Ibid.

[85] General Secretariat of the Council 'Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - Member States Comments on Public Sector and Chapter IX' (2014) <https://data.consilium.europa.eu/doc/document/ST-14098-2014-REV-1/en/pdf>.

[86] Ibid.

from the permissive approach to data processing facilitated by this amendment. The contribution of research to society is not expressed in economic terms but rather its ability to obtain new knowledge of great value. Recital 125 notes the possibility of social science capitalising on advances in data processing technologies, enabling researchers to 'obtain essential knowledge about the long-term impact of a number of social conditions e.g., unemployment, education, and the coupling of this information to other life conditions.' There is an explicit link between the activity of research and tangible impacts in references to its ability to

> *provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people, and improve the efficiency of social services etc.*[87]

This exclusive focus on the positives of research results in its construction as an activity in the public interest that ignores the potential for commercial entities to undertake such activities. Furthermore, this shapes an understanding of 'societal expectations' that would be permissive of such data reuses due to their benefits and the apparent lack of negative consequences or risks. The Council of the EU was not the only institution relevant to the drafting process of legislation in the EU. The following section will introduce the European Parliament and explore its role in the creation of the GDPR.

2.3.3.2 The Committee System Within the European Parliament

As the only directly elected institution with legislative powers, many describe the European Parliament as the 'voice' of EU citizens.[88] Feliu and Serra posit that as

[87] Ibid.

[88] Laura Feliu and Francesc Serra, 'The European Union as a 'Normative Power' and the Normative Voice of the European Parliament' in Stelios Stavridis and Daniela Irrera (eds), The European Parliament and its International Relations (Routledge 2015).

the European Parliament is 'elected in direct representation of citizens,' it is in a 'leading position' to protect individuals.[89] This reliance on public elections means that Members of the European Parliament (MEPs) are more likely to be sensitive to human rights concerns and issues that become salient within public discourses.[90] This means that the European Parliament both promotes human rights' protections and applies pressure on other EU institutions to 'achieve greater consistency in the defence of universal human rights.'[91] It is not just an awareness of the electorate's opinions that promoted more rights-based framings of the proposed legislation. The practices of the committees within the European Parliament influenced the construction of key terms and approaches within the GDPR.[92]

The lead committee for the legislative file of the GDPR was the Committee on Civil Liberties, Justice and Home Affairs (LIBE.) It was tasked with drafting the reports on the progress made within the European Parliament whilst other committees merely submitted their opinions on the proposals. This role accorded the LIBE committee a greater level of power over the legislative process than other committees. It facilitated its ability to promote certain discourses that aligned with its views concerning the appropriate approach to legislative reform. The LIBE committee managed the 'file' of the GDPR, increasing its influence over the

[89] Ibid, 20.

[90] Ibid.

[91] Ibid.

[92] Matthieu Moulonguet, 'What Drives the European Parliament? The Case of the General Data Protection Regulation' [2016] Bruges Political Research Papers No 47.

legislative process by making it easier to include its proposed amendments compared to MEPs within other committees.[93] It also had an informational advantage over other committees due to the comparative ease with which it could track the proposal's progress. It was thus seen as the entity best placed to amend the file,[94] further increasing its power to shape and dominate discourses. The view of Thomas Van der Valk, parliamentary assistant of MEP Sophia In't'Veld demonstrates the expertise attached to the committee. He felt that

> *if you have the Industry committee or the Internal Market committee deciding rules that have such deep impacts on the personal lives of citizens, then you will not have the people with the right skills to work on it.*[95]

The choice of Rapporteur reinforced the understanding of the LIBE committee as the entity best placed to amend and track the progress of the proposal. The Rapporteur is responsible for steering a proposal through Parliament during the legislative process. Some argue the role 'enables individual legislators to take responsibility for, and exert influence over, the policy position of the Parliament as a whole.'[96] During the drafting of the GDPR, the Rapporteur for the LIBE committee was Jan Philipp Albrecht,[97] a member of the Green group and well known for his strong views on privacy and data protection. Albrecht's

---

[93] Ibid.

[94] Ibid.

[95] Ibid.

[96] Pierre Hausemer, 'Participation and Political Competition in Committee Report Allocation' (2006) 7 European Union Politics 505, 509.

[97] Wikipedia Contributors, 'Jan Philipp Albrecht' (2021) <https://en.wikipedia.org/wiki/Jan_Philipp_Albrecht>.

legal background and substantial engagement with the issues arising from the reform attempts led to him having a substantial influence over the approach taken by the European Parliament.

Within its proposals, the LIBE Committee included a direct reference to Member States' positive obligations under the European Convention on Human Rights (ECHR) to ensure the appropriate regulation of data flows. This constructs a problematisation that seeks to respond to rights-based issues by increasing the human rights material, emphasising the importance of human rights in this domain. The reference to the ECHR supports a problematisation rooted in rights-based concerns, as the Convention is not a human rights instrument of the EU's own making. Instead, it is the product of a different institution and has its own enforcement mechanisms.[98] This makes it less likely to be interpreted by its judges to support the EU's Internal Market objectives. For this thesis, however, the main changes created by a rights-framed problematisation were the significant constraints upon data processing for research purposes. It rejected the idea that the processing of sensitive data for historical, statistical, and scientific research purposes was as urgent as public health and social protection. Therefore, social research did not warrant exceptions that would place it 'on the same level' as the other listed justifications. Instead, the Committee proposed that processing personal data for research purposes should rely on the data subject's consent, except where the Member States provided exceptions in cases that served 'an

[98] European Court of Human Rights, 'European Court of Human Rights' (2021) <https://www.echr.coe.int/Pages/home.aspx?p=home>.

exceptionally high public interest.'[99] The approach was defended upon the basis that, without recourse to the need to demonstrate an exceptionally high public interest, 'any "research", no matter if academic or corporate and including e.g., market research, could be used as an excuse to override all protections provided for in the other parts of this Regulation.'[100]

However, the final version of the GDPR does not reflect this rights-based framing of the data protection reform, despite external events demonstrating the importance of rights-based protections in this context. What led to the minimisation of rights-based discourses in favour of the EU's internal market objectives? The final section explores why these developments did not transform into a sustained shift towards a rights-based approach to the GDPR.

2.3.3.3 The Need for an Agreement Between the EU Parliament and the Council of the EU

The ordinary legislative procedure, a bicameral approach that requires both the European Parliament and the Council of the European Union to approve the proposed legislation, produced the GDPR. Once the Commission proposes a piece of legislation, the respective Parliamentary Committees and Council consider these proposals independently yet parallel. A proposal can go through two 'readings,' where the institutions can accept the other's proposal or propose

[99] Committee on Civil Liberties, Justice and Home Affairs, 'Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))' (2013) <https://www.europarl.europa.eu/doceo/document/LIBE-PR-501927_EN.pdf?redirect>.

[100] Ibid, 203.

amendments. A failure to achieve a compromise during the second reading triggers the creation of a conciliation committee in a final attempt to find a proposal amenable to both. Where a conciliation committee fails to find an acceptable compromise, the Council and the Parliament retain the power to block the proposal.[101]

The GDPR required a second reading but passed without recourse to a conciliation committee. As the negotiations dragged on, it became clear that both institutions desired a swift resolution. In 2015, the Council of the European Union reached an agreement on a general approach to the legislative proposal. This gave the 2015 Luxembourg presidency a negotiating mandate to begin trilogues with the European Parliament. The trilogues saw the proposal of numerous research-friendly amendments, including the presumed compatibility of further processing of personal data for research purposes. The broad definition of research, encompassing commercial entities, meant these amendments served both academic and commercial purposes. Therefore, these proposals still contributed to the market-making aims, even if such an aim was not explicitly considered. The presidency also sought to introduce a Recital recognising that 'it is not possible to fully identify the purpose of data processing for scientific purposes at the time of data collection.'[102] Based on this, the ability for a data subject to 'give their consent

[101] Craig and De Búrca (n 70).

[102] The Presidency of the Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)' (2015) <https://data.consilium.europa.eu/doc/document/ST-10391-2015-INIT/en/pdf>.

to certain areas of scientific research when in keeping with recognised ethical standards for scientific research'[103] was also proposed.

Different discourses are prominent in the respective approaches of the Council and Parliament. The Parliament sought to permit derogations and exemptions to reconcile the right to personal data protection with the rules governing the freedom of expression, including academic expression. The Council disagreed with this approach, suggesting it did not support the assumption that these two rights must be traded off against one another. Nevertheless, it proposed a new Recital that still emphasised the non-absolute nature of the right to the protection of personal data. Not only must it be 'balanced' with other fundamental rights, including ones that promote economic interests, but it must also regarding its 'function in society.' Recitals are an interesting feature of EU lawmaking. They do not constitute laws themselves despite being part of the same text. Instead, they operate to outline the 'worldview'[104] of the EU on the subject, further contributing to the attempts to construct key terms and shape ways of thinking. The orientation of the proposal indicates its function was to promote a free flow of data and support the economic aims of the EU. The inclusion of several rights the GDPR proposed to 'respect' and 'observe,' including the freedom to conduct a business, further supports this view. During the latter stages of negotiations, differences between the institutions were evident in their respective approaches towards the derogations applying to data processing for research purposes.

[103] Ibid.

[104] Faulkner (n 12).

In contrast to the Parliament's approach, the Council perceived such processing as appropriate even if non-identifiable data could fulfil the data processing purpose and the 'highest technical standards' were not used. One possible explanation for this difference is that the Council's proposals adopt the risk-based approach that received substantial support from the Member States' delegates. An alternative explanation is that whilst both actors utilised a risk-based approach, the Parliament's proposals indicate a view of research as an inherently risky activity, which justified the imposition of higher technical standards and limitations. This demonstrates the potential impact of different problem representations constructed by the respective institutions. The Council emphasised the proposals to promote the Digital Single Market and the EU as the world leader in processing data to facilitate its economic activities.

The Parliament's cautious approach focussed on the new intrusions these developments permitted into individuals' interests and sought to constrain them accordingly. Furthermore, the Parliament's approach was not just more restrictive; it was also more prescriptive. It included several measures to 'minimise the processing in compliance with the proportionality and necessity principles.'[105] This aligns with Ripoll Servent's[106] analysis on the differences between the institutions, whereby the European Parliament started with notions

[105] European Parliament, 'European Parliament Legislative Resolution of 12 March 2014 on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))' (2014) <https://www.europarl.europa.eu/doceo/document/TA-7-2014-0212_EN.html?redirect>.

[106] Ariadna Ripoll Servent, 'Protecting or Processing? Recasting EU Data Protection Norms' in W Schünemann and MO Baumann (eds), *Privacy, Data Protection and Cybersecurity in Europe* (Springer 2017).

of ‘liberty’ within a ‘security’ context. The more centrist and right-wing parties challenged this view by promoting an understanding of liberty underpinned by an economic rationale. The latter rationale came to dominate through its support in the Council’s approach, focussed on facilitating data flows that promoted the economic opportunities fostered by data processing.

Nevertheless, an agreement was reached, and the GDPR entered into force in May 2016, before becoming directly applicable in all Member States in May 2018. But what were the effects of the problematisation that created the final GDPR text? The next chapter explores this question, moving from the drafting process to consider the effects of the legislation through the framings and discourses present in the final version.

3 EXPLORING THE EFFECTS OF THE GENERAL DATA PROTECTION REGULATION

3.1 INTRODUCTION

The following builds on Chapter 2, considering the effects engendered by the discourses that were present before and during the legislative process. The term 'effects' rather than 'outcomes' is deliberate in light of Bacchi's argument that the former covers a wider range of impacts that may be 'much more subtle in their influence.'[1]

This chapter firstly explores the effects of the General Data Protection Regulation[2] (GDPR) in terms of its 'silences.' This acknowledges what remains unsaid or unthinkable due to the discourses, practices, and actors that shaped the legislative solution may nevertheless have effects. The GDPR's discursive effects are analysed to explore how discourses shaped and influenced what was deemed permissible, and its impact on the legal governance of reusing administrative data in social science research. This section presents four 'truths' generated by the discourses that shaped the legislative text. The following section considers the subjectification effects of the GDPR, drawing upon Foucauldian ideas as to how human beings are 'made' subjects through their interactions with forms of power, including the law.[3] The section focuses on the effects of the GDPR in 'making' a

[1] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

[2] 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.

[3] Hubert Dreyfus and Paul Rabinow, *Michel Foucault: Beyond Structuralism and Hermeneutics* (Routledge 2016).

certain type of human being as a 'data subject.' The final section considers the lived effects of the GDPR, exploring what the GDPR has changed, who has been harmed because of its provisions, and who benefits.

3.2 Silences that Promote Reuses of Data in Research

One set of effects generated by the GDPR arise from aspects of data processing that were explicitly or implicitly deemed 'unproblematic,' therefore existing as a 'silence' within the law. This section analyses the constraints upon the problematisation fostered by such silences.[4] Three 'silences' within the GDPR of particular relevance are the continuation of a 'technology-neutral' approach to data protection law, the impact of certain constitutional developments within the EU, and the relationship between the activity of research and the academic community. These silences, and the assumptions underpinning their supposedly unproblematic nature, are highly salient to exploring potential divergences between the legal approach and societal expectations towards reusing data in research. They provide further insight into the construction of key terms and approaches within the GDPR, the underlying assumptions of the legislative approach, and the subsequent effects.

One 'silence' is the continued applicability of the technology-neutral approach towards data processing. This approach was present in the previous 1995 Data Protection Directive[5] and attempts to prevent unforeseen advances in

[4] Bacchi (n 1).

[5] EU Directive 95/46: 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.

technology rendering the law obsolete by not being based on any particular technology. Yet, if the Directive is problematised as outdated and unable to respond to the modern data processing landscape, this undermines the appropriateness of this approach. Nevertheless, the GDPR explicitly adopts a technology-neutral approach, casting aspersions on the ability of the legislative reform to respond to novel challenges. This implies that other drivers may underpin the calls for legislative reform.

The lack of explicit recognition of certain constitutional developments is a further 'silence' within the legislative process. The first development was the introduction of Article 16 of the Treaty on the Functioning of the European Union (TFEU), which created a new standalone basis for EU legislation in the field of data processing and the free movement of data. The second is the introduction of the Charter of Fundamental Rights,[6] which was accorded the same status as the Treaties of the EU in the 2009 Treaty of Lisbon. The Commission Communication of 2012 suggests the inclusion of Article 16 was understood, or at least constructed as, imposing an obligation or duty upon the EU to act. The Commission emphasises that '[d]ata protection is a fundamental right in Europe, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union, as well as in Article 16(1) TFEU, and needs to be protected accordingly.'[7] The European Data Protection Supervisor, commenting on the Commission's Communication, noted

[6] EU Charter of Fundamental Rights: Charter of Fundamental Rights of the European Union, OJ 2010 C 83/389.

[7] Commission, 'Safeguarding Privacy in a Connected World a European Data Protection Framework for the 21st Century (Communication)' (COM(2012) 9 final, 25 January 2012), 2.

the Lisbon Treaty marked 'a new era' for personal data protection through this direct legal basis.[8] The Impact Assessment of the draft GDPR echoes this sentiment, linking the Treaty of Lisbon to 'why it is time to build a stronger and more coherent data protection framework in the EU.'[9] The previous Directive was based upon Article 114 TFEU, permitting the adoption of measures that seek to establish the proper functioning of the internal market. Due to subsequent developments in the EU, Article 16 TFEU means data protection is now a *lex specialis*, emancipated from internal market objectives and the protection of the right to privacy. Hijmans[10] argues the values listed in Article 2 TEU, which references the 'values of respect for human dignity, freedom, democracy, equality, the rule of law, and respect for human rights' as the foundational values of the EU. Consequently, such values should infuse all activities undertaken in the EU and are fundamental to determining the appropriateness of actions taken by the EU. Therefore, though Article 16 TFEU is a *lex specialis* for data protection, Hijmans suggests the 'centre of the gravity' of data protection has changed to give more importance to its fundamental rights facet.[11]

However, the analysis within the chapter indicates that the discourses around the appropriate uses of data do not necessarily reflect this shift. Instead,

[8] European Data Protection Supervisor, 'A Comprehensive Approach on Personal Data Protection in the European Union (Opinion)' (2011).

[9] Commission, 'Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) (Impact Assessment)' (2012).

[10] Hielke Hijmans, *The European Union as Guardian of Internet Privacy: The Story of Art 16 TFEU* (Springer 2016).

[11] Ibid.

the market framing still dominates the interpretation of the rights-based protections. This framing suggests that whilst these new provisions were acknowledged as the 'trigger,' the silence as to the difference this would make to legislation based upon them enabled the market-framing approach to dominate still. The European Data Protection Supervisor recognised that a

> *strong framework for data protection is the necessary consequence of the importance given to data protection under the Lisbon Treaty, particularly Article 8 Charter of the Fundamental Rights of the Union and Article 16 TFEU.*[12]

They also stated, however, that

> *a strong framework for data protection also serves wider public and private interests in an information society with ubiquitous data processing. Data protection fosters trust, and trust is an essential component of the well-functioning of our society. It is essential that arrangements for data protection are construed in a way that they — as much as possible — actively support rather than hamper other legitimate rights and interests.*[13]

Therefore, despite the increased prominence of the rights-based facets of data protection, data is still framed through a market discourse where 'knowledge is the engine of [the EU] economy. And data is its fuel.'[14] The best ways to regulate modern data processing practices are thus those that 'maximise the value and minimise the cost of data.'[15] In the Impact Assessment of the draft GDPR, the focus remained on developing rules that 'will allow the digital economy to develop

[12] European Data Protection Supervisor (n 8), 6.

[13] Ibid.

[14] Neelie Kroes, 'The Economic and Social Benefits of Big Data' (Webcast Conference on Big Data 2013).

[15] Ibid.

across the internal market' without a specific reference as to how rights will develop as part of a 'stronger and more coherent data protection framework.'[16] There is also silence regarding how this will facilitate a more permissive approach and how it appears to undermine citizen desires for further control. Alongside these silences, the discursive effects of what was said are pertinent to the analysis, which the following section explores.

A further silence in the discourses that constructed the GDPR is how far the research is no longer within the exclusive domain of academia. The 2020 Opinion published by the European Data Protection Supervisor[17] highlights this point. In contrast to the silence on this matter during the creation of the GDPR, the Opinion notes that the

> *boundary between private sector research and traditional academic research is blurrier than ever, and it is ever harder to distinguish research with generalisable benefits for society from that which primarily serves private interests.*[18]

This emphasises the silences present during the GDPR's creation that helped prevent a discussion that could have constructed a 'common understanding of which activities qualify as genuine research.'[19] The Opinion's exploration of the 'intertwining of academia and the commercial sector' brings into sharp relief the silence present during the creation of the GDPR regarding the potential influence

[16] Commission (n 9).

[17] European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (2020) <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>.

[18] Ibid, 2.

[19] Ibid.

of commercial entities upon research. In minimising the discourses that explore the interplay between commercial entities and research, this silence minimises alternative views that may construct research as an activity not justifying the permissive governance approach of the GDPR. Alongside these silences, the discursive effects of what was said are pertinent to the analysis and are explored in the following section.

3.3 Discursive Effects that Promote Reuses of Data in Research

This analysis of the 'discursive effects' draws upon Bacchi's ideas about the necessity of analysing the 'limits imposed on what can be thought and said'[20] to understand a text better. Discursive effects explore how the identified problem representations and the discourses that frame them 'make it difficult to think differently, limiting the kinds of social analysis that can be produced.'[21] Language does not operate to describe an objective reality. Instead, it creates the social worlds that people experience[22] and, by constructing the issues at hand, it exposes the 'central questions of truth and power.'[23] These 'socially produced forms of knowledge' then become 'accepted truths.'[24] The imposition of limits and promotion of certain discourses may harm some and benefit others. Four 'truths'

[20] Bacchi (n 1).

[21] Ibid.

[22] Murray Edelman, *Constructing the Political Spectacle* (The University of Chicago Press 1988).

[23] Frank Fischer and John Forester, *The Argumentative Turn in Policy Analysis and Planning* (Duke University Press 1993).

[24] Carol Bacchi and Susan Goodwin, *Poststructural Policy Analysis a Guide to Practice* (Palgrave Macmillan 2016).

promulgated in the dominant discourses within the GDPR, which are of relevance to the reuse of data in research, will be discussed in this section. They are the framing of the developments as an opportunity to be seized, the 'truth' that research warrants a more permissive approach towards regulating its data processing activities, the appropriateness of a risk-based response, and the appropriateness of retaining certain principles present in the 1995 Directive.

3.3.1 The 'Truth' is that Developments Provide an Opportunity that Must be Seized.

The most significant innovations in data processing practices concern the advances made possible by big data analytics and the increasingly sophisticated analyses performed upon big data datasets.[25] This includes the expectation that big data would be a tool for evidence-informed policy and decision-making, enabling administrations to make the 'right' choices more quickly.[26] However, the ability to use insights generated by big data relies upon many other factors, undermining the certainty of such an outcome. Furthermore, whilst the outcomes of research using big data analytics require a coherent data ecosystem, a commitment to the reuse of government data, and investment in the training and education of both data scientists and policymakers,[27] the GDPR does not regulate

[25] Gernot Rieder, 'Tracing Big Data Imaginaries through Public Policy: The Case of the European Commission' in Ann Rudinow Sætnan, Ingrid Schneider and Nicola Green (eds), *The Politics of Big Brother: Big Data, Big Brother?* (Taylor Francis 2018).

[26] Ibid.

[27] Gunther Oettinger, 'A Digital Single Market: The Key to Europe's Industrial Leadership in the Digital Economy' (ICT2015, 2015).

these factors. Nevertheless, the advances made possible by big data justify the EU's actions as it 'can't miss out on that kind of growth opportunity.'[28]

Those beliefs underpin discourses that promote the need for the EU to capitalise on these advances so that it can be 'at the forefront of the global data race,'[29] justifying the GDPR's approach. Within the text, the discourses concerning research make it challenging to construct research as a potentially harmful activity. Doing so closes off opportunities to consider the myriad of ways research can affect individuals' lives, which may potentially warrant a more nuanced, contextually dependent approach to its regulation. This becomes incorporated into the GDPR's construction of 'reasonable expectations, the 'public interest' and the 'function in society,' which shapes the operation of individuals' rights.

Whilst these discourses justify a permissive approach to reusing data in research, it is unclear whether individuals share these underlying assumptions and justifications. The understanding of the definition of legitimacy used within this thesis considers if the actions of an entity are deemed proper, appropriate, or desirable by reference to a system of socially constructed norms, values, and beliefs.[30] However, the ways of thinking that are closed off by this silence risk the law being justified by reference to a system of values and beliefs not shared by

[28] Neelie Kroes, 'The Big Data Revolution' (EIT Foundation Annual Innovation Forum, 2013).

[29] Commission, 'European Commission and Data Industry Launch €2.5 Billion Partnership to Master Big Data (Press Release)' (2014) <https://ec.europa.eu/commission/presscorner/detail/en/IP_14_1129>.

[30] Mark Suchman, 'Managing Legitimacy: Strategic and Institutional Approaches' (1995) 20 Academy of Management 571.

those in society affected by the law. The next section will discuss a further silence relevant to the construction of research within the GDPR.

3.3.2 The 'Truth' is that Research Warrants a More Permissive Approach

The discourse in the GDPR constructs research homogenously, as exemplified by the broad definition of scientific research in Recital 126. This states that

> *[s]cientific research for the purposes of this Regulation should include fundamental research, applied research, and privately funded research and in addition should take into account the Union's objective under Article 179(1) of the Treaty on the Functioning of the European Union of achieving a European Research Area.*[31]

The discursive effect imbues all such research with the legitimacy traditionally afforded to certain types of research, particularly non-commercial research. The EDRi[32] proposed to amend this Recital to align the definition of scientific research with Article 13 of the EU Charter. This states that '[t]he arts and scientific research shall be free of constraint. Academic freedom shall be respected.' [33] This amendment, however, was not supported. Therefore, an opportunity to align the GDPR with the Charter was rejected in favour of an approach that facilitated a permissive approach to data processing. Consequently, the GDPR is founded upon the 'common assumption that scientific research is

[31] Recital 126, 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.

[32] EDRi, 'Proposed Amendments to Recital 126' (*EDRi's comments on the Data Protection Reform,* April 16, 2012) <https://protectmydata.eu/recitals/recitals-121-139/recital-126/>.

[33] Charter of Fundamental Rights of the European Union, OJ 2010 C 83/389.

beneficial to the whole of society and that scientific knowledge is a public good to be encouraged and supported.'[34]

Further Recitals portray research as an activity in the public interest. The effect is to frame research as such an inherently 'good' thing, making it hard to consider its potential to cause harm and limiting certain forms of social analysis to this effect.[35] The discursive effects construct research as an indispensable activity to modern society as it is the main means for ensuring an effective response to societal challenges. Recital 126, in its assertion that research leads to 'essential knowledge,' may be a way of undermining discourses that have a more cautious attitude towards the outcomes of research. Despite the inherently uncertain nature of research, the discourse within the GDPR presents the benefits as certain. This is a further example of the epistemological consequences of big data analytics, where the Recitals indicate a reification of the future that crystallises a reality presented as infallible.[36] By presenting this future as the only possible one, this discourse narrows down the options for alternatives. Such options could be other ways to respond to societal challenges or the inability of research to address these issues successfully. As stated previously, the silences

[34] European Data Protection Supervisor (n 17).

[35] See Recital 157 GDPR which states '*By coupling information from registries, researchers can obtain new knowledge of great value with regard to widespread medical conditions such as cardiovascular disease, cancer and depression. On the basis of registries, research results can be enhanced, as they draw on a larger population. Within social science, research on the basis of registries enables researchers to obtain essential knowledge about the long-term correlation of a number of social conditions such as unemployment and education with other life conditions. Research results obtained through registries provide solid, high-quality knowledge which can provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people and improve the efficiency of social services.*'

[36] Stefania Milan and Lonneke van der Velden, 'The Alternative Epistemologies of Data Activism' (2016) 2 Digital Culture & Society.

may lead to certain norms, values, and beliefs of the society governed by the GDPR not being reflected in the law or considered during the drafting process. The presence of these silences may therefore contribute to an understanding of those relying upon the law to justify their data processing activities as not acting in a way deemed proper, appropriate, or desirable by certain citizens.

The exceptional approach justified by the research outcomes is present in the framing of research in Recital 33. It permits using broad consent for scientific research purposes when

> *in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.*[37]

The references to 'recognised ethical standards' suggest that research uniquely engages the interests that are protected and promoted by data processing rules. Consequently, it implies that the rules outside the realm of data protection better serve to reconcile the engaged interests. Doing so limits the ability to consider some highly relevant questions that analyse the underlying justifications and question the implicit 'truth' underpinning this statement. Instead, this discourse minimises the attention directed to exploring what it is about these standards that justify individuals' exercising less control over their data. It also overshadows discourses that question how and why they appropriately protect individuals' interests in a way that the rights-based protections present in the GDPR do not. Therefore, this silence minimises the impact of the norms, values, and beliefs that underpin this discourse. This creates the potential for lawful

[37] Recital 33 GDPR.

actions to be perceived as inappropriate or undesirable by certain citizens who hold the views that align with the silenced discourses. The lack of a specific reference to a document outlining these 'recognised standards,' means one can only speculate what effect these standards could have on individuals that would justify them having less control over their data.

Furthermore, the reference to these standards promotes a discourse of research as an activity that is governed, actively and knowingly, by high ethical standards that ensure the appropriate protection of individuals. Allowing researchers to govern their actions suggests a high level of respect for researchers and trust in their ability to govern research practices appropriately. It suggests a level of homogeneity and infallibility is inherently present within these guidelines as a means of ensuring the appropriate protection of data subjects. It is not clear whether this truly aligns with the reality of ethical standards utilised in research settings. Instead, this may result from more dominant discourses shaping and limiting the kinds of analysis and 'truths' present in the GDPR.

A further discursive effect results from including the EU's own human rights instruments and advocating its own institutions and rules as the most suitable for protecting individuals in this context. Such 'limits' on the relevance and merit of other instruments are supported by the failed attempts of the Committee on Civil Liberties, Justice and Home Affairs (LIBE) to include direct references to Member States' positive obligations under the European Convention on Human Rights (ECHR.) The aim was to 'ensure that such data flows are appropriately regulated' and that fundamental rights were not 'undermined' by

the GDPR.[38] Its absence in the final version indicates the dominant discourse did not perceive this instrument as necessary to ensure the appropriate protections.

An alternative explanation is that actors involved in promoting the dominant discourses may have preferred the approach of the Charter as a comparatively less restrictive set of protections. The prominent discourse present in the GDPR aligns with ideas proposed by other Committees that focus on the EU's non-rights-based objectives. For example, the Committee on the Internal Market and Consumer Protection's Opinion proposed to include the need for 'a proper balance between [the] protection of privacy and respect of the single market.'[39] In doing so, it specifically cautioned against allowing data protection rules to 'undermine competitiveness, innovation, and new technology,'[40] with no corresponding references to the rights-based implications of a data protection framework reform. The discursive effects that promote this truth are also present in specific sections of the GDPR, which will be explored in turn below.

3.3.2.1 Article 89 GDPR

Article 89 governs the data processing for 'archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' and

[38] Committee on Civil Liberties, Justice and Home Affairs, 'Amendments 351-601 on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)' (2013) <https://www.europarl.europa.eu/doceo/document/LIBE-AM-504340_EN.pdf?redirect>, 5.

[39] Committee on the Internal Market and Consumer Protection, 'Opinion on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)' (2013) <https://www.europarl.europa.eu/doceo/document/IMCO-AD-496497_EN.pdf?redirect>, 5.

[40] Ibid.

permits certain derogations from individuals' rights. Data processing for research purposes enables the Member States to provide for derogations from the rights of access, rectification, restriction of processing, and objection. Despite strong public saliency over the risks posed by innovative data processing activities and doubts expressed within the LIBE Committee's proposals regarding the value of research, the final text demonstrates a very permissive approach towards using and reusing data for research purposes.

By grouping research in this way, the discursive effects embodied within Article 89 close off the ability to consider whether certain instances of processing for research purposes should benefit from these exemptions and derogations. This suggests that the actions deemed desirable and appropriate are based on the view that a broad understanding of research should benefit from the approach of Article 89. Furthermore, it inhibits discussion regarding the scope for individuals to determine whether they believe such instances of data processing justifies significant derogations from, and limitations to, their rights. Despite an apparent emphasis on putting the individual back in control of their data, the discursive effect of Article 89 seeks to exclude data processing for research purposes from this data subject-centred approach, considering the public benefits arising from such uses of data. This aligns with the discourses highlighted by Starkbaum and Felt, who saw the push from actors from the data-dependent fields of scientific research shift 'the discourse from its key topic of data protection,' to a paradigmatic focus on 'protecting the health of individuals and societies at large.'[41]

[41] Johannes Starkbaum and Ulrike Felt, 'Negotiating the Reuse of Health-Data: Research, Big Data, and the European General Data Protection Regulation' (2019) 6 Big Data & Society 1, 2.

These discourses link to the broader trend of depicting big data approaches as necessary innovations that serve the public good. This minimises the scope for considering the risks such approaches may engender. Starkbaum and Felt[42] argue that the common discourses deployed by these actors, aligned in their 'argumentative repertoires,' facilitated the discursive effect of permitting this permissive approach towards data processing for research purposes. This was achieved by constraining explicit discussions on the issues or emphasising the broader societal benefits, so these disadvantages were successfully 'balanced out.' This links to arguments that the definition of risk is highly contingent upon the 'temporal, socio-political and institutional context,' leading it to be a 'fluid construct.'[43] The meaning of risk, therefore, can be mediated and reconfigured to rationalise a range of actions.[44] In this context, such circumstances make it very difficult to challenge and oppose the prevailing discourses regarding the benefits of research referenced in Article 89 due to the framing of its wider benefits. This framing thus justifies intrusions into individuals' interests. Doing so risks the resulting law being deemed illegitimate by failing to fully reflect the system of norms, beliefs, and values held by those governed by the GDPR. This may lead to the action of those following the law being deemed inappropriate or undesirable. Nevertheless, it remains unclear how far the construction and understanding of

[42] Ibid.

[43] Kelly Hannah-Moffat, 'Algorithmic Risk Governance: Big Data Analytics, Race and Information Activism in Criminal Justice Debates' (2018) 23 Theoretical Criminology 453.

[44] Ibid.

research promoted within the GDPR align with societal views towards research. Do individuals share this expansive notion of the benefits?

3.3.2.2 Article 85 GDPR

Article 85 governs the processing of data whereby there is a need 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.*[45]

It permits a range of derogations by providing a new opportunity to tailor the provisions to data processing for academic purposes, as distinguished from the historical and scientific purposes covered by Article 89. Its inclusion was partially justified by practical concerns identified by the Wellcome Trust that 'research in areas such as politics and history is unlikely to be compatible with the research model set out [Article 85 GDPR] and may not be permitted otherwise.'[46] Moreover, the ESRC noted the need to bring academic social science research into the ambit of this provision to ensure that a 'historian or social investigator working in an academic context should not be treated less favourably by the law than [one] writing in a non-academic context.'[47]

This innovation has the discursive effect of placing academic data processing on the same 'level' as other rights engaged in this context, including

[45] Article 85(1) GDPR.

[46] Beth Thompson, 'Academic Research Perspective on the European Commission, Parliament and Council Texts of the Proposal for a General Data Protection Regulation' (2015) <https://wellcome.org/sites/default/files/research-perspective-data-protecton-regulation-proposal-wellcome-jul15.pdf'>.

[47] Economic & Social Research Council, 'Response to the European Commission's Proposed European Data Protection Regulation (COM (2012) 11 Final)' (2013)

those protecting individuals' rights. This aligns the activity of research with the right to freedom of expression. This right to freedom of expression is protected due to its ability to enable societal development and progression,[48] facilitate self-fulfilment, and protect individual liberty.[49] Mill wrote extensively on how freedom of expression was vital to the attainment of knowledge and the search for truth, arguing that

> *[i]f the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error.*[50]

Freedom of expression further enables individuals to inform themselves about matters of public concern, providing a potential check against government overreach.[51] Erdos[52] justifies the inclusion of academic expression by referencing normative reasons concerning why social science research should benefit from the permissive approach where data processing falls within the scope of freedom of expression and falls within the scope of the GDPR. He presents an image of research that is

> *oriented to making public information and ideas that creates high-value outputs, far more worthy than some of the journalistic*

---

[48] Josie Timms, 'Why Is Free Speech Important?' (2016) <https://www.indexoncensorship.org/2016/04/free-speech-important>.

[49] Brittan Heller and Joris Van Hoboken, 'Freedom of Expression: A Comparative Summary of United States and European Law' (2019) <https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/Freedom_of_Expression_TWG_Heller_May_2019.pdf>.

[50] John Stuart Mill, *On Liberty.* (Arcturus Publishing 1859).

[51] Heller and Hoboken (n 49).

[52] David Erdos, 'Reconciling Humanities and Social Science Research with Data Protection' (2016) <https://slideshare.net/DavidErdos1/reconciling-humanities-and-social-science-research-with-data-protection>.

> *'infotainment' that also permit derogations to be made from data protection law.*[53]

This chimes with Dingwall's point that

> *[i]f what has traditionally been the most disinterested source of information, the universities, become systematically handicapped in that competition [between information providers], then all citizens lose out.*[54]

By including academic data processing within the domain of activities protected by the right to freedom of expression, the GDPR has the effect of reinforcing the notion that such data processing serves both societal and individuals' interests. This discursive effect justifies the derogations from the rights found within the GDPR.[55] Its inclusion further reinforces the expansive definition of research present within the GDPR, ensuring the broadest range of data processing activities can benefit from the derogations and exemptions made possible by its provisions. This promotes a construction of research as an activity that is equally worthy of the protections offered by Article 89 GDPR, despite empirical evidence suggesting that individuals do not necessarily see the value of social or historical research.[56] The frequent highlighting and emphasising of social

---

[53] Ibid.

[54] Robert Dingwall, 'The Ethical Case against Ethical Regulation in Humanities and Social Science Research' (2008) 3 Twenty-First Century Society 1, 10.

[55] British Academy and Economic & Social Research Council, 'British Academy and ESRC Press for Shields for Humanities and Social Science Scholarship as UK Implements the New EU General Data Protection Regulation' (2017) <https://www.thebritishacademy.ac.uk/documents/103/2017.06_-_BritAc__ESRC_GDPR_Submission_summary_3.pdf>.

[56] Daniel Cameron, Sarah Pope and Michael Clemence, 'Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes' (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/> accessed February 21, 2021.

science research as a 'publicly-focussed activity'[57] distances academic data processing from the potential negative consequences of innovative data processing. Instead, it is aligned with activities historically protected through rights that protect freedom of expression. The discursive effect promotes an image of academic outputs being beneficial for both societies and individuals.

The inclusion of academic processing within this section facilitates this use of data by creating an 'extremely strong default'[58] in favour of its permissibility. The justification for this approach is that 'people have a right to disseminate public knowledge…but also go through all the steps preparatory which are necessary to that, collecting the data, analysing the data, and of course, receiving the data.'[59] The discursive effect constructs an image of academic outputs as shaped by the 'ethics of culmination, precision, and reflexivity' that serve individual and societal interests by producing high-value materials.[60] This further emphasises the reasonableness of these derogations.[61] The verb to 'reconcile' creates a discursive effect of an outcome whereby tensions between the competing rights at play can

[57] British Academy and Economic & Social Research Council, 'A British Academy and Economic & Social Research Council Submission to the Department of Culture, Media and Sport (DCMS) Call for Views on the General Data Protection Regulation Derogations' (2017) <https://www.thebritishacademy.ac.uk/documents/104/2017-05-10_DCMS_GDPR_Derogations_submission.pdf>.

[58] Full Fabric, 'The Impact of the GDPR in Academic Research - a Talk by Dr. David Erdos' (2017) <https://blog.fullfabric.com/the-impact-of-the-gdpr-in-academic-research-a-talk-by-dr.-david-erdos>.

[59] Ibid.

[60] Erdos (n 52).

[61] Economic & Social Research Council, 'Response to the European Commission's Proposed European Data Protection Regulation (COM (2012) 11 Final)' (2013) <https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-response-to-the-european-commission-s-proposed-european-data-protection-regulation-2013/>.

be resolved to facilitate their protection and promotion. This limits the ability to consider the potential for tension and clashes between rights to occur. Furthermore, the notion of reconciliation is a very neutral term that may belie some of the difficult decisions that the Member States must make when considering how to promote the rights to the protection of personal data and freedom of expression within their respective legal systems. By emphasising the freedom of information, one can see the approach of Article 85 as emphasising the importance of an output and, as information may be commodified, the internal market framing discourse prominent through the GDPR.

This discursive effect minimises the scope for considering how certain characteristics of academic data processing are not inherent. Instead, these characteristics arise through carefully planned, considered, and evaluated research plans that remain responsive to potential ethical issues. Despite this, the GDPR leaves undefined the central concept of 'academic expression.' The British Academy and ESRC refer to them as a 'relatively small number of potential derogations' when one considers and recognises 'research's essential role in knowledge production of all kinds,'[62] they undermine most of the protections available to individuals. This is particularly challenging when one considers the extent to which individual control dominates the discourse surrounding individuals' rights. If the Member States exercise these derogations, many of the requirements to process data lawfully, including the need to have a lawful basis for data reuse, could be overruled. The transparency obligations, one of the

[62] British Academy and Economic & Social Research Council (n 55).

GDPR's cornerstones, could also be negated—along with the need to ensure accuracy and data minimisation.[63] What remains unclear is how far these discursive effects reflect the system of norms, beliefs, and values shared by those affected by the GDPR and how far the discursive effects may lead to the actions of data processors being deemed inappropriate or undesirable.

3.3.3 The 'Truth' is that Individual Rights Provide Sufficient Protections

Like the previous Directive, the GDPR relies on individual rights to protect individuals from unwarranted intrusions or exclusions due to data processing activities. Whilst the basis for such protections has shifted from the right to privacy to the right to the protection of personal data, they still share the same underlying assumption. They both assume the protection of rights is rooted in individual control.[64] The discursive effect thus constructs the strengthening of rights as a positive step. Individual rights are a means of empowering individuals to take control of their data and uphold and promote their autonomy.[65] This view is present in the literature[66] and case law.[67] The discursive effect of its inclusion presents this continuation as a positive and empowering mechanism for

[63] J Bell and others, 'Balancing Data Subjects' Rights and Public Interest Research: Examining the Interplay between UK Law, EU Human Rights Law and the GDPR" (2019) 5 European Data Protection Law Review 43.

[64] See, for example, Alan Westin, *Privacy and Freedom* (Ig Publishing 2015) and Samuel D Warren and Louis D Brandeis, 'The Right to Privacy' (1890) 4 Harvard Law Review 193.

[65] Adam Moore, *Privacy Right: Moral and Legal Foundations.* (Pennsylvania State Univ 2011).

[66] James Whitman, 'The Two Western Cultures of Privacy: Dignity versus Liberty' (2003) 113 Yale Law Journal 1151.

[67] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014]

individuals to use. Nevertheless, this glosses over how modern data processing practices led to a significant power asymmetry between data controllers and data subjects.[68]

The 'truth' presented by this approach towards protecting individuals' interests belies the criticisms that the complexities of modern data processing activities undermine an individual's ability to make an informed choice[69] and to uphold their rights effectively.[70] This discursive effect also glosses over one of the significant shifts triggered by innovative data processing practices. Modern data processing is less focussed on individuals but instead seeks to investigate the attitudes and behaviours of larger groups and communities.[71] Floridi argues that as re-identifiable groups are targetable groups, it is thus 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.'[72] Innovations mean that individuals can be clustered without needing to identify them and traditional identifiers are increasingly irrelevant to the purposes for processing data. These 'proxies' for

[68] Alessandro Mantelero, 'The Future of Consumer Data Protection in the E.U. Re-Thinking the 'Notice and Consent' Paradigm in the New Era of Predictive Analytics' (2014) 30 Computer Law & Security Review 643.

[69] Julie Cohen, 'Turning Privacy inside Out' (2019) 20 Theoretical Inquiries in Law 1.

[70] Jef Ausloos and Pierre Dewitte, 'Shattering One-Way Mirrors – Data Subject Access Rights in Practice' (2018) 8 International Data Privacy Law 4.

[71] Mantelero (n 68).

[72] Luciano Floridi, 'Open Data, Data Protection, and Group Privacy' (2014) 27 Philosophy & Technology 1.

attributes that are protected by law are difficult to predict and protect, particularly in the context of big data analytics.[73]

Furthermore, rights may be given, but if they are perceived as confusing or burdensome then individuals will be unable or unwilling to assert them. This renders any strengthening of rights in this context a nullity.[74] Empirical research suggests individuals approve of the rights bestowed upon them by data protection law but are uncertain about how to go about exercising them.[75] The sense of control fostered by the language of individual rights implies an ability to effectively participate in the data processing activities using individuals' data. It also presumes the ability to make informed and voluntary choices about how their data should be processed.[76] However, as already noted in the context of group rights and collective interests, an external entity, such as a data controller, often creates these groups. The individuals within these ad hoc groups, therefore, lack the intentionality and self-awareness to protest.[77] This inability to prevent harm indicates a problematisation which led to a solution that encourages novel uses of data and new analyses, even at the expense of individuals' interests.

---

[73] Tal Zarsky, 'Transparent Predictions' (2013) 4 University of Illinois Law Review 1503.

[74] ICO, 'Data Protection Rights What the Public Want and What the Public Want from Data Protection Authorities' (2015) <https://ico.org.uk/media/about-the-ico/documents/1431717/data-protection-rights-what-the-public-want-and-what-the-public-want-from-data-protection-authorities.pdf>.

[75] Cameron and others (n 56).

[76] Christophe Lazaro and Daniel Le Métayer, 'Control over Personal Data: True Remedy or Fairytale?' (2015) 12 SCRIPTed.

[77] Brent Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' (2017) 30 Philosophy & Technology 475.

3.3.4 The 'Truth' is That Existing Principles Still Fulfil Their Function Within Data Protection Law

The inherently innovative nature of research using big data analytics poses fundamental challenges to many of the core principles of data protection law, especially the principle of data minimisation.[78] Rubinstein[79] finds the principle 'inimical' to the nature of big data analytics. They argue that ex-ante restrictions thwart its ability to discover new correlations through applying sophisticated analytic techniques to massive data sets. The continued reliance on such principles as providing the foundation for data protection law, even after legislative reform is in 'stark tension,' with the fluid and norm-challenging nature of many types of social research.[80] Tene and Polonetsky[81] argue that data minimisation is not a 'practical' approach to privacy in the 'age of big data.' Instead, this interest must be 'balanced against additional societal values such as public health, national security and law enforcement, environmental protection, and economic efficiency' in a way the principle of data minimisation fails to facilitate. They argue that the continued focus on data minimisation, 'with little appreciation for the value of uses for data, could jeopardise innovation and beneficial societal advances.'[82] The emphasis on protecting privacy fostered by the principle of data minimisation fails

[78] David Erdos, 'Systematically Handicapped? Social Research in the Data Protection Framework' (2011) 20 Information & Communications Technology Law 83.

[79] Ira Rubinstein, 'Big Data: The End of Privacy or a New Beginning?' (2013) 3 International Data Privacy Law 74.

[80] Erdos (n 78).

[81] Omer Tene and Jules Polentsky, 'Privacy in the Age of Big Data a Time for Big Decisions' (2012) 64 Stanford Law Review.

[82] Ibid.

to account for other values. Zarsky raises a similar point that more data should theoretically generate greater knowledge 'and thus greater benefit to the firms and potentially society in general.'[83] The discursive effect of this principle reduces concerns about the mere holding of personal data being sufficient to undermine individual autonomy and generate anxiety.[84] This would seem at odds with the market framing approach of the GDPR. However, Mayer-Schönberger and Padova[85] note that many examples of big data analytics will benefit from the exceptions and derogations within the GDPR, where data are processed and reused for research purposes. Research purposes are not narrowly defined in the legislation and could incorporate a wide range of activities.

The principle of anonymisation determines the scope of data protection law. However, big data analytics undermine the de-identification techniques that attempt to anonymise data.[86] The notions of 'identified' and 'identifiable' remain crucial to distinguish the type of data processed and, consequently, the application of the GDPR.[87] The discursive effect of this distinction constructs the processing of anonymous data as an activity that will not negatively affect an individual's interests, despite this undermining of many de-identification techniques. [88] Furthermore, anonymity is a technique used to prevent any potential negative

[83] Tal Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 Seton Hall Law Review.

[84] Ibid.

[85] Viktor Mayer-Schönberger and Yann Padova, 'Regime Change: Enabling Big Data through Europe's New Data Protection Regulation' (2016) 17 Columbia Science and Technology Law Review 315.

[86] Mantelero (n 68).

[87] See Recital 26 GDPR.

[88] Ibid.

impacts on an individual. This does not prevent an individual from being harmed through data processing that facilitates group profiling. Despite the discursive effect of constructing anonymisation as the appropriate means to determine boundaries of what data processing requires legal governance and what does not, a different justification for its inclusion may be inferred. Anonymisation under the GDPR, upon closer inspection, appears to be less about ensuring the law protects data that could cause individuals harm were it to be misused and more about facilitating the free flow of data.

There are elements of a 'relative' instead of an 'absolute' approach to identifiability within the GDPR, which suggests a more permissive approach to data processing.[89] Adopting an 'absolute' approach would mean data would be identifiable if even one person possesses the key that could decode the data. This approach towards identifiability would entail a definition of 'personal data' that encompasses a significant amount of data and would be broader than the Directive's approach. This could potentially hinder data processing where it was previously considered anonymous, as it would now have to comply with the requirements of the GDPR. In contrast, a 'relative' approach considers the effort in terms of time, money, and expertise required to identify the data subject and would require more than a highly theoretical possibility of identification for data to be considered personal. Spindler and Schmechel[90] note that the definition of

[89] Miranda 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 222.

[90] Gerald Spindler and Philipp Schmechel, 'Personal Data and Encryption in the European General Data Protection Regulation' (2016) 7 Journal of Intellectual Property, Information Technology and E-Commerce Law 163.

personal data requires one to consider the means 'reasonably likely to be used' that, through the notion of 'reasonableness,' is indicative of a relative approach. The Article 29 Working Party's interpretation of identifiability, based on the Data Protection Directive, supports this view. It held that 'a mere hypothetical possibility to single out the individual is not enough to consider the person as 'identifiable.'[91] The preference for a relative approach is also supported when one considers that the right to the protection of personal data is not absolute but, instead, is to be considered in relation to its function in society. Its function is to facilitate the use of data whilst ensuring the appropriate protection of individuals' rights and freedoms. Therefore, this shift to a risk-based approach may be a further example of the market framing of data protection legislation.

Discursive effects are only one type of effect relevant to assessing the construction of the appropriate reuses of data in research within the GDPR. The following section analyses the subjectification effects that analyse how discourses foster certain understandings of the individuals affected by the legal reform.

3.4 SUBJECTIFICATION EFFECTS

The idea that discourses within laws and policies make certain 'subject positions available'[92] draws upon the work of Foucault on subjectification, exploring how human beings are 'made' subjects.[93] Of importance to this thesis is that human

[91] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (WP 136, 20 June 2007).

[92] Bacchi (n 1).

[93] Dreyfus and Rabinow (n 3).

self-understanding, the 'truth' about who we are, is a result of interactions with forms of power. One form of power is the law. Therefore, where individuals assume the subject position made available by the discourses present in laws and policies, they tend to make 'sense of the social world from this standpoint.'[94]

3.4.1 The Data Subject is Informed and Able to Exercise Control

Provisions of the GDPR reveal that it constructs data subjects as sufficiently rational and informed to use the tools it provides to promote and protect their interests. Individuals are constructed as rational actors, capable of weighing the value of their personal information against the benefits of its wider use, and knowledgeable as to when and how their data may be used. This effect aligns with a key justification for the GDPR: the need to facilitate individuals' ability to exercise control over their data. The proposals made by the LIBE Committee during the drafting process of the GDPR reflect this:

> *The right to the protection of personal data is based on the right of the data subject to exert the control over the personal data that are being processed. To this end the data subject should be granted clear and unambiguous rights to the provision of transparent, clear and easily understandable information regarding the processing of his or her personal data, the right of access, rectification and erasure of their personal data, the right to data portability and the right to object to profiling...The provisions of this Regulation should strengthen, clarify, guarantee and where appropriate, codify those rights.*[95]

LIBE's calls to require data subjects' consent to permit the processing of personal data concerning health necessary for historical, statistical, or scientific

[94] Bacchi (n 1).

[95] Committee on Civil Liberties, Justice and Home Affairs, 'Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))' (2013).

research purposes[96] further supports this view. The Committee on the Internal Market and Consumer Protection's Opinion echoes this emphasis on individual control and infers that such control could be effective. It constructs data protection by design as a 'useful tool' due to its ability to allow 'the data subject to be fully in control of his own data protection.'[97]

There is a silence in the GDPR as to the burden this places on individuals. Instead, the subjectification effects reinforce the notion that 'full control' is a positive thing for individuals to wield. This reflects a broader trend towards an increased focus on the individual in EU policymaking.[98] The EU has increasingly constructed individuals as rational *homo economicus*, who can protect their interests where they receive sufficient information. The GDPR constructs the principle of transparency as improving individuals' understanding and awareness of data processing practices involving their data and promoting their effective control over their personal data. This has the subjectification effect of constituting alert and informed data subjects, yet it is not clear this subjectification effect reflects the reality. As argued by Edwards and Veale,[99] this subject position requires 'constant vigilance,' and it is 'unreasonable and increasingly unsustainable' to expect data subjects to manage their interests. Individuals are

[96] Ibid.

[97] Committee on the Internal Market and Consumer Protection, 'Opinion on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)' (2013).

[98] Bart van der Sloot, 'Do Data Protection Rules Protect the Individual and Should They? An Assessment of the Proposed General Data Protection Regulation' (2014) 4 International Data Privacy Law 307.

[99] Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a Right to Explanation Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18.

'mostly too time-poor, resource-poor, and lacking in the necessary expertise'[100] to find and understand information about reuses of their data, even if attempts at transparency are made. The innovations fostered by big data analytics exacerbate these issues by undermining individuals' ability to comprehend transparency options and their awareness of when their rights can be used.[101]

Rather than acknowledge the burden placed on individuals, the problematisation constructed the rights bestowed by the GDPR as empowering individuals.[102] Fougère and others argue that empowerment may be a signifier of responsibilisation as it entails a partial shift in responsibility to individuals and a behavioural change towards more 'responsible individual practices in relation to different societal challenges.' [103] Including an individual's 'reasonable expectations' creates a subject with such expectations, a proposition not entirely supported by empirical research.[104] This effect facilitates the construction of a homogenous set of data subjects and places great power in the hands of data processors. Where expectations are not clear, what points of reference are data processors to use? They may rely on the subject constructed within the GDPR, who supports research and can utilise their rights where they feel their interests are infringed. In doing so, they seek to speak 'in the name of people,' silencing those

[100] Ibid.

[101] Simone Fischer-Hubner, 'Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work?' *10th IFIP International Conference on Trust Management* (2016).

[102] Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014].

[103] Martin Fougère, Beata Segercrantz and Hannele Seeck, 'A Critical Reading of the European Union's Social Innovation Policy Discourse: (Re)Legitimizing Neoliberalism' (2017) 24 Organization 819.

[104] Cameron and others (n 56).

represented and minimising their ability to voice and address differences in opinion.

The subjectification effects present in the dominant discourses in the GDPR thus 'silence' alternative reasons for focusing on individuals. For example, the Council of Europe instead linked this focus to the 'banalisation of data processing rather than an attempt to strengthen individuals' rights.[105] As data processing has become increasingly intrinsic to the needs of private companies and not just the public sector, the general rules of fairness and legitimacy may not fit as well with the modern data processing landscape. Furthermore, these rules may not be understood and employed in the same way by different actors.

The risk-based approach impliedly constructs a data subject that supports research by placing further power in the hands of data processors. Data processors are empowered to determine the stringency of the protections to be offered, and data subjects are constructed as tacitly supporting a level of 'risk' to their rights. The subjectification effect of the right to be informed is that an individual is assumed to understand such information. Nevertheless, the provision of information does not necessarily correlate with an increase in understanding of the processes involved in certain data processing activities.[106] Furthermore, it is also a controversial inclusion as it relies on responsible parties to identify risks

[105] Council of Europe, 'New Technologies: A Challenge to Privacy Protection?' (1989) <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000001680684607>.

[106] P van de Waerdt, 'Information Asymmetries: Recognizing the Limits of the GDPR on the Data-Driven Market' (2020) 38 Computer Law & Security Review 105436.

and act accordingly without any interference from the State. [107] The subjectification effect of this is that individuals' failures to live up to these standards are constructed as resulting from their rational and behavioural defects. [108] This construction of the data subject aligns with the influence of neoliberalism where the 'extreme emphasis on individualism can be seen to lead to the depoliticization of social problems through transforming them into individual problems with market solutions.'[109] This resonates with neoliberalism which disseminates market rationality to all spheres of human action. The effect on the individual is to reframe human action as 'rational entrepreneurial action driven by a cost-benefit analysis.'[110] The neoliberal subject is further constructed as one that accepts the omnipresence of risk and are 'responsibilised' for their possible failure because an 'assessment of risk should inform their utility-maximizing calculation.'[111]

3.4.2 The Data Subject is Supportive of Research.

A further subjectification effect is the construction of individuals as supportive of research. This support necessarily encompasses the range of exemptions and derogations available where data are processed for research purposes. The most

[107] Lina Jasmontaite and others, 'Canvas White Paper 2 – Cybersecurity and Law' (2017) <https://dx.doi.org/10.2139/ssrn.3091939>.

[108] Tobias Matzner and others, 'Do-It-Yourself Data Protection—Empowerment or Burden?' in Serge Gutwirth, Ronald Leenes and Paul De Hert (eds), *Data Protection on the Move - Current Developments in ICT and Privacy/Data Protection* (Springer 2016).

[109] Fougère, Segercrantz and Seeck (n 103).

[110] Ibid.

[111] Ibid.

overt example of this subjectification effect is present in Recital 113 of the GDPR that states:

> *For the purposes of processing for historical, statistical and scientific research purposes, the legitimate expectations of society for an increase of knowledge should be taken into consideration.*[112]

The subjectification effect legitimises research as beneficial and as a data processing activity that ought to occur. Despite contradictory empirical data, the law is not constructed as imposing a view upon society but as responding to societal views.[113] It also has the subjectification effect of creating a homogenous society, precluding space for debate and disagreement. Instead, the positives of research and its ability to improve people's lives are emphasised without acknowledging the difficulties in translating research into beneficial outcomes for individuals and society. This Recital, combined with the subjectification effects fostered by the GDPR, guides the construction of the right to the protection of personal data considering its function in society. Therefore, the right should not prevent this legitimate expectation of society and be an impediment to research. This interpretation facilitates research, prompting a reconcilement of the interests in data protection law that promotes the use of data, justified through creating a subject that expects this use of their data.

3.4.3 The Data Subject as a Rights-holder in an Improved Position

The GDPR constructs individuals as proactive rights holders, able to use this increased level of control over their data to make informed choices. However, this

[112] Recital 113 GDPR.

[113] Cameron and others (n 56).

construction is, in many circumstances, a façade. The innovations in modern data processing practices exacerbate the recognised issues of information deficits and asymmetries.[114] The explicit focus on transparency within the GDPR will not necessarily aid individuals in overcoming these issues as its success is based on two presumptions. Firstly, a reliance on transparency presumes that individuals have the time, inclination, and ability to discover what data are collected about them and learn how they are used. However, the significant increase in passive data collection undermines the ability of individuals to know what data are collected about them. Furthermore, learning about how data are used may require a level of specialist knowledge that a substantial number of individuals lack. Secondly, reliance on transparency is flawed in its presumption that the collection, linkage, analysis, and storage processes undertaken as part of big data processing can be communicated to enable individuals to understand without rendering the explanation so simplistic as to be untruthful. The utility of control is premised on the individual's ability to weigh the costs and benefits of a particular data processing activity and exert their choices in a meaningful and effective manner. It is questionable how far this premise is present in the context of reusing administrative data in social science research due to its reliance on passive data collection and relative unfamiliarity to many in society.

[114] Gordon Hull, 'Successful Failure: What Foucault Can Teach Us about Privacy Self-Management in a World of Facebook and Big Data' (2015) 17 Ethics and Information Technology 89.

3.5 LIVED EFFECTS

The lived effects explored in this section refer to how the dominant discourse and representations in the GDPR impact people's lives.[115] This facilitates an exploration of the changes; what has stayed the same, who may be harmed, and who may benefit from the legislative reform.[116] Here, individuals appear to have more rights but less control, remaining susceptible to the risks created by novel data processing practices that include reuses of data in research. The discursive and subjectification effects of the GDPR place the onus on individuals, constructed as knowledgeable, rational, and capable beings, to know and uphold their rights. The lived effects of the GDPR chime with Golder's work that considers Foucault's later interest in human rights. According to this analysis, rights are

> *one of a range of contingent political tools available for counter-investment and appropriation, for "strategic reusability" on behalf of different political interests and as a part of diverse political struggles.*[117]

The lived effects stem from a discourse defined by Bergemann[118] as the 'data market world' that constructs data protection as serving the 'higher goal of creating mutual benefits.' This further supports the construction of subjects as consumers, emphasising the notion of individuals playing an active role, despite empirical evidence suggesting the contrary.[119] Therefore, where data are used for

---

[115] Bacchi (n 1).

[116] Ibid.

[117] Ben Golder, *Foucault and the Politics of Rights* (Stanford University Press, 2015), 22.

[118] Benjamin Bergemann, 'The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection' in Marit Hansen and Others (eds), *Privacy and Identity Management The Smart Revolution* (Springer 2018).

[119] See, for example, Joanna Strycharz, Jef Ausloos and Natali Helberger, 'Data Protection or Data Frustration? Individual Perceptions and Attitudes towards the GDPR' (2020) 6 European

a purpose that individuals are unaware of and potentially not supportive of, the lived effects foster a sense of blameworthiness and a failing on their part. It infers that those individuals have failed to appropriately employ the tools and information available to assert their rights properly. Individuals must bear this burden with little acknowledgement of the issues created by power asymmetries between data subjects and processors, and individuals' dependence on services that demand their data. This facilitates the attribution of responsibility to be sidestepped by other actors who could address the structural problems that engender such effects.[120]

This lived effect chimes with Foucault's reflections on governmentality and the notion of 'responsibilisation.' This is a neoliberal mode of governance that renders subjects individually responsible for tasks that may not have been previously recognised as their responsibility.[121] Where the choice of options are equated with individual choice and an expression of free will, the consequences are borne by the data subject alone.[122] Hull[123] recognises that norms and rules regarding privacy have distributional consequences and that a focus on the individuals' needs and wants subverts the social benefits of privacy. The societal benefits are difficult to quantify, yet it is down to the individuals to choose how

Data Protection Law Review 407 and Nora A. Draper, 'From Privacy Pragmatist to Privacy Resigned: Challenging Narratives of Rational Choice in Digital Privacy Debates' (2016) 9 Policy & Internet 232.

[120] Matzner and others (n 108).

[121] Ibid.

[122] Thomas Lemke, 'The Birth of Bio-Politics': Michel Foucault's Lecture at the Collège de France on Neo-Liberal Governmentality' (2001) 30 Economy and Society 190.

[123] Hull (n 114).

best to assert their rights, a widely recognised feature of neoliberal governmentality.[124]

3.6 CONCLUSIONS

This chapter sought to demonstrate that what happened before and during the legislative process that culminated in the GDPR prioritised and promoted certain discourses. The aim was to provide a grounding to compare the legal approach to reusing data in research with societal expectations. The constructivist approach facilitated an unpacking of the influences on the drafting process. The analysis provided insights into the construction of key terms and mechanisms for justifying data reuses, highlighting the dominance of market-framing discourses. This raises questions about how far legislative terms that seek to protect individuals and ground the permissibility of data use on individuals' support and expectations diverge from societal expectations. However, the GDPR was not the only legislative response to the developments identified in Chapter 1, and the following chapter explores a different legislative response.

[124] Ibid.

4 EXPLORING A LEGAL CONSTRUCTION OF THE PERMISSIBILITY OF THE REUSE OF DATA IN RESEARCH: THE DIGITAL ECONOMY ACT 2017

4.1 BACKGROUND

The Digital Economy Act 2017 (DEA) was introduced to Parliament in July 2016 and received Royal Assent in April 2017.[1] In contrast to the General Data Protection Regulation (GDPR), the scope of the DEA goes beyond data processing and data protection to legislate on access to digital services, digital infrastructure, and intellectual property.[2] The placing of data processing in this broader context provides a potentially different approach towards determining what secondary uses of administrative data in research could be legally permissible. Unlike the GDPR, the DEA focuses much less on modernising legislative approaches to data processing by reference to the need to protect better individuals' rights. This focus suggests that a different problematisation shaped the legislative response of the DEA. Furthermore, the process by which the DEA became legislation warrants further analysis. The legislative process engaged a wide range of stakeholders, collecting input in various ways to shape the problematisation justifying the DEA. An analysis of the Open Policy Making process (OPM process), public consultations, and expert witnesses during the Bill's scrutiny by the Public Bills Committee, a parliamentary committee that scrutinises proposed bills and

[1] UK Parliament, 'Digital Economy Act 2017 – Stages' (2017) <https://bills.parliament.uk/bills/1859/stages>.

[2] See the Introduction to the Digital Economy Act 2017.

suggests amendments, could provide insights into how certain actors and power relations shaped the DEA.

The first section explores the impact of the initial triggers for the DEA. It highlights how discourses and practices during the legislative process problematised the existing legal landscape as engendering unjustified barriers. This was problematised as preventing the UK from capitalising on the opportunities presented by a range of technological advances. The previous chapter acknowledged that multiple forms of knowledge produce legislation, which is reflected in the practices that created the DEA. Its initial triggers and subsequent developments incorporated various participation practices that included experts and citizens. The literature concerning the development of laws that govern emerging technologies notes that expert involvement is a strategy to 'shape' lawmaking to deal with the uncertainties generated by these advances.[3] The perceived knowledge of experts provides legitimacy through enhancing the perceived neutrality of the legislative outcome. [4] Considering this thesis' understanding of legitimacy, this may strengthen the assumption that the governor's actions are proper through using more expert and neutral actors. Therefore, this broad participation enlarges the democratic legitimacy of such legislation by better enabling norms, values, and beliefs outside the political domain to influence the legislative process.[5] The chapter then analyses the DEA's

[3] Alex Faulkner and Lonneke Poort, 'Stretching and Challenging the Boundaries of Law: Varieties of Knowledge in Biotechnologies Regulation' (2017) 55 Minerva 209.

[4] Ibid.

[5] Ibid.

evolution, exploring the underlying presumptions and broader discourses that justified certain approaches. Finally, it investigates the effects generated by the DEA through the interplay of the different factors and discourses that shaped its approach as embodied in the final text. The focus of this chapter is on Part 5 of the DEA, as the aspect most relevant to secondary uses of administrative data in research.

4.2 Generating the Need for the Digital Economy Act 2017

This analytical approach foregrounds the socially constructed nature of problems that laws intend to 'solve.' Laws do not respond to objectively verifiable issues that await discovery. Instead, they result from practices, discourses, actions, inactions, and compromises between the needs and desires of many actors with differing levels of power. This section explores the impact of three prompts for legislative reform. The first is a set of discourses that suggested a need for legal reform, based on the UK starting to 'fall behind' other nations that permitted greater reuses of administrative data. This provoked a response based on the idea that the UK should strive to be a world leader in research, justifying a prompt and wide-ranging legislative response. The second is a discourse justifying reform based on societal expectations towards the use of government-collected data. As a trigger, it legitimises the decision to intervene in this area and suggests that the driver for the choices made were citizens' expectations. This implies that societal norms and views permeate throughout the legislative process. The final trigger is the discourses that suggest that failures to capitalise on the opportunities presented

by greater reuses of administrative data negatively impact economic and societal interests.

4.2.1 Not Letting the UK 'Fall Behind'

A trigger for the DEA was a set of discourses that portrayed the UK as falling far behind other nations in its ability to capitalise upon technological advances. As a result, the UK was constructed as unable to use these advances to gain new insights into societal challenges that could form the basis of new legislation and policies. This problematisation, defined as the specific socially constructed problem to which the DEA became the 'solution,' promoted certain discourses. The discourses perceived a need to overcome the apparently 'complex legal landscape.'[6] The discourse further emphasised the opportunities rendered possible by legislative reform. These opportunities included 'the delivery of better public services through world leading research and statistics'[7] and the ability to 'improve the lives of citizens,' including the increased use of administrative data for research purposes.[8]

These discourses, promulgated by different bodies within the government, serve to legitimise attempts to promote a more permissible approach to data use by highlighting other countries that have already adapted to this way of using data. The discursive effect undermines the persuasiveness of opposing discourses

[6] Department for Culture Media & Sport, 'Digital Economy Bill Factsheet - Digital Government: Research and Statistics (Clauses 56-68)' (2016) <https://www.gov.uk/government/publications/digital-economy-bill-part-5-digital-government>.

[7] Ibid.

[8] Cabinet Office, 'Better Use of Data in Government - Consultation Paper' (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government>.

that seek to promote a more cautious approach. This potentially challenges how far society's norms, values, and beliefs, by which the governor's actions are assumed to be desirable and appropriate, are reflected in the DEA. The dominant discourse minimises opposing discourses by constructing these discourses as against the UK becoming a world leader or denying the best for UK citizens. Such views, even if only inferred, are contrary to the objectives of those involved in political and lawmaking processes. References to the actions of other nations were not the only trigger, and the next section will consider a more 'domestic' trigger.

4.2.2 Ensuring the Law Meets Societal Expectations.

Another trigger for the DEA was the 'Better Use of Data in Government'[9] public consultation that took place between February and April 2016 and received 282 responses. Practices such as public consultations seek to justify the outcomes of the legislative process as providing outcomes that consider a wide range of views, implying a level of societal support for the result. It provides an opportunity to render more explicit the socially constructed set of norms, values, and beliefs used to determine how far the lawmakers' actions may be assumed to be appropriate and desirable. The 'Better Use of Data in Government' public consultation document promotes a discourse that links the need for reform to societal expectations. It claims that the UK government has

> *a duty to provide public services to help those with the least, building a society where there is real opportunity for all. To do this, we need to keep pace with both rising public expectations and the availability of new technology.*[10]

[9] Ibid.

[10] Ibid, 1.

The DEA Explanatory Notes echoes this by stating the 'public increasingly expects to access services quickly and conveniently, at times and in ways that suit them.'[11] These discourses heighten the legitimacy and perceived democratic credentials of any resulting legislative reform by suggesting the values and beliefs society were considered and included. They provide an interesting contrast to the discourses present in the shaping of the GDPR. The rights-based discourses that played a role in the GDPR are barely present here, promoting constructions of key terms and approaches that further promote permissive approaches to reuses of administrative data in research.

These discourses dominated alternate discourses that expressed a more cautious understanding of societal expectations. Evidence submitted by the Administrative Data Research Network[12] (ADRN) acknowledged potential public concern yet found a more permissive approach could 'gain public trust easily' if done in a certain way. This does not entirely contradict the above, but it did emphasise the need for a 'publicly acceptable means of securing the massive societal, governmental, and service benefits that [data] linkages can offer.'[13] The focus on the significant benefits further contributes to the discourses that prompted a particular response through the DEA, which shaped the understanding of societal expectations which helped justify the legislative outcome as the 'solution.' Societal interests were constructed as sufficiently

[11] Explanatory Notes to the Digital Economy Act 2017, para 27.

[12] Digital Economy Bill Deb, 20 October 2016 vol 615, col 246W.

[13] Ibid.

pressing to necessitate legislative reform in ways beyond references to societal expectations. The following section explores how discourses concerning the negative impacts of the current legal landscape on both economic and societal interests triggered calls for new legislation that culminated in the DEA.

4.2.3 Negative Impacts on Economic and Societal Interests

The discourses that highlighted the wide range of interests potentially harmed by inaction fostered a stronger case for reform by appealing to many stakeholders. The apparent consensus of the support for reform encouraged by this discourse legitimises the legislative 'solution' by appearing to consider all those impacted. This renders it difficult for other actors to oppose its approach. Discourses surrounding the potentially jeopardised interests shape specific constructions of key terms such as 'science' and 'research' that may influence the approach taken during the legislative process.

Within the 2015 Conservative Manifesto,[14] 'science' is considered within the auspices of 'industrial strategies' and 'enterprise.' Such discourses promote an understanding of research that values its outcomes in terms of its contribution to economic interests, rather than for knowledge's sake. Whilst briefly acknowledged as 'worthwhile in its own right,' the indisputable focus within the Manifesto is how science can ensure UK entities can profit from their research and support industrial activities. Science underpins much of the investment proposed by the Conservatives as a means for providing jobs, supporting a 'strong economy to help

[14] The Conservative Party, 'The Conservative Party Manifesto 2015' (2015) <https://issuu.com/conservativeparty/docs/ge_manifesto_low_res_bdecb3a47a0faf>.

you and your family' and making Britain the 'best place in Europe to innovate.'[15] This desire to be innovative is linked to the aim to 'be number one in Europe and in the top five worldwide in the World Bank's Doing Business rankings by 2020.'[16] Science is viewed through the prism of enterprise, where support is offered to ensure universities can 'make money from the technology they develop' and direct 'further resources towards the Eight Great Technologies…where Britain is set to be a global leader.' The dominant discourse constructs the benefits of research as tangible, financial, and focussed on more short-term gains. This is critical as the Manifesto underpinned the legislative efforts of the Conservative government, including the passing of the DEA.

The evidence collected before the legislative process chimes with the Manifesto's dominant discourses. The ADRN's submission began with the economic benefits fostered by potential reforms, with precise figures, before moving to a much briefer outline of the less precise 'wider social and economic impact [which] could include better health outcomes, improved economic growth.'[17] The identification of economic benefits in these discourses facilitates a permissive approach towards data processing. Doing so elides the domains of science and economy, shaping the former's aims by reference to the latter's yardstick. This blurring between the two shapes the DEA's legislative approach to promote and facilitate the type of science and research that produces economic benefits. The discourses emphasising the potential loss of economic benefits as

[15] Ibid, 21.

[16] Ibid, 17.

[17] Digital Economy Bill Deb (n 12).

detrimental to wider society support this approach. The UK government's discourse promotes a construction of the 'public benefit' in this context that values more quantifiable goals and 'efficient' services rather than the protection of individuals' rights. It is unclear how far these values dominate the socially constructed values that assist in assessing the desirability and appropriateness of action permitted by the DEA. Before exploring this further empirically, one must explore how the discourses discussed in this section developed during the legislative process. The following section, therefore, analyses the interplay of discourses and practices during the development of the DEA.

4.3 DEVELOPMENTS

This section analyses some of the discourses and practices that took place during the development of the DEA. This explores the problematisation that prompted a legislative response and how the DEA became the 'solution.' The key milestones in the DEA's development that are considered include the Open Policy Making process, the Better Use of Data in Government Consultation, the Impact Assessment for the Digital Economy Bill, and the process by which the Digital Economy Bill passed into law.

It is not just these milestones and events that are relevant to understanding the DEA's development. The broader context in which the UK government operates emphasised certain discourses that influenced the DEA's development. The accompanying Explanatory Notes align the provisions of the DEA with other

initiatives, including the 2012 Government Digital Strategy.[18] This outlined how the government was to become 'digital by default' in response to the perceived wishes of the public to access public services conveniently and quickly. The Strategy posited that data linkage, which could facilitate insights into citizens' habits, behaviours, and needs, would support these efforts. Concurrently with the development of the DEA, the government consulted upon and published the Government Transformation Strategy[19], which outlined the process by which it intended to develop into a digital organisation. This process included making a 'better use of data,' a phrase echoed in the consultation paper that influenced the legislative process of the DEA. The focus of these practices is far removed from the rights-based, control-oriented understandings of permissible data use that some empirical studies have suggested, particularly in the context of health research.[20] Nevertheless, a claim to citizen support is still justifiable by reference to such discourses in the Conservative Party Manifesto.[21] This enables a claim to public support for these actions, supporting the legitimacy of the resulting law, despite the lack of specific detail regarding facilitating reuses of administrative data in research within the Manifesto.

---

[18] Cabinet Office, 'Government Digital Strategy' (2012) <https://www.gov.uk/government/publications/government-digital-strategy>.

[19] Government Digital Service, 'Government Transformation Strategy' (2017) <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy>.

[20] BS Buckley, AW Murphy and AE MacFarlane, 'Public Attitudes to the Use in Research of Personal Health Information from General Practitioners' Records: A Survey of the Irish General Public' (2010) 37 Journal of Medical Ethics 50.

[21] The Conservative Party (n 14).

4.3.1 The Open Policy-Making Process

The Open Policy Making process (OPM process) shaped the discourses found within the DEA, which helped construct key terms and its approach towards reuses of data. This ran from 2014 to 2016, facilitating collaboration between central government departments, local authorities, and civil society organisations. It aimed to explore the potential for legislative change to enable reuses of administrative data. The inclusion of a broad range of stakeholders bestows a greater legitimacy upon the resulting legislation by undermining claims that the DEA solely served the UK government's aims. This practice supports the idea of a political acceptance of enabling what works based on experts' views rather than what suits their agenda. As stated by Francis Maude, the Minister for the Cabinet Office during the process;

> *experience from across government had led to the conclusion that not removing the barriers to data sharing was shortchanging the public. However, the Government had learnt the lessons from previous attempts to pass legislation and understood that the process of developing the policy needed to be done in a different way. It needed to be genuinely open and would only be successful if it achieves a high degree of consensus. The process should be open to all organisations with knowledgable interest at every stage – sharing and listening to concerns and working things out in a genuinely collaborative way.*[22]

The membership of the OPM process meant its knowledge base was technical, but its purpose was to go beyond such boundaries. The novel linkages fostered by this process is an example of what Faulkner and Poort describe as an amplified entanglement of technical and regulatory knowledges within political

[22] Involve, 'Planning the Open Policy Process' (2014) <https://web.archive.org/web/20160814104446/http://datasharing.org.uk/2014/04/03/planning-the-open-policy-process-31-mar-2014-meeting-note/>.

debates involving political strategy.[23] The highlighting these actors' expertise and neutrality heightens the legitimacy of their outputs, but the practice remains inherently political and subject to factors deemed important by that sphere of activity.

During the OPM process, participants emphasised the value of transparently running these sessions to enable public discussion.[24] Even if attempts at such discussions were unsuccessful, such calls for discussion and transparency would legitimise the resulting discourses by emphasising the process's inclusive nature. Participants' efforts to determine the best way to engage with the public to understand their beliefs reinforce this view. They requested more details of the proposals to enable the public to understand them, focussing on using case studies to demonstrate potential benefits.[25] This practice grounds the proposals in the 'reality' faced by data sharing entities and highlighted the importance of public understanding as a requisite for the necessary public support.

Participants complained of certain actions, or inactions, as undermining their ability to contribute and minimising the impact of their feedback and guidance. This de-emphasis on the competing dimensions of the problem definition within these more 'inclusive' practices conceals how the OPM process was an

[23] Faulkner and Poort (n 3).

[24] Involve (n 22).

[25] Involve, 'De-Identified Data for Research | 13 June 2014 | Meeting Note - Data Sharing' (2014) <https://web.archive.org/web/20161109012339/http://datasharing.org.uk/2014/06/18/de-identified-data-for-research-13-june-2014-meeting-note/>.

instrument of power wielded by the government.[26] Therefore, despite the process apparent inclusivity and the emphasis on attaining a 'better,' more balanced, legislative solution, much of the DEA aligns with governmental priorities and objectives. This suggests that the values underlying the government's actions may not be perceived as desirable or appropriate by citizens.

Whilst elements of the OPM process were carried forward during the developments of the DEA, a 'rushed' publication of the Bill minimised its impact. The Bill was published despite a 'general lack of clarity, particularly on safeguards, and many questions were still being asked, such as how, why, when and so on.'[27] Notwithstanding the appearance of a practice that was designed to ensure the law 'did not raise hackles, that it got the privacy and safeguards right,'[28] participants felt 'it lacked the safeguards everybody had been saying were needed.'[29] These comments shed light on how the presence of a practice can justify a solution, even if its participants felt it might not have realised its potential for fostering broader public engagement. The drafting of the DEA incorporated another practice that sought to ensure a more inclusive legislative process. This next section explores the impact of the public consultation on constructing key terms and prioritising certain discourses and knowledges.

[26] Thomas Archibald, 'What's the Problem Represented to Be? Problem Definition Critique as a Tool for Evaluative Thinking' (2019) 41 American Journal of Evaluation 6.

[27] HC Deb 11 October 2016, vol 615 col 61.

[28] Ibid.

[29] Ibid.

4.3.2 The Better Use of Data in Government Consultation

Though no specific consultation took place on the DEA, the UK government ran a public consultation on the 'Better Use of Data in Government.'[30] This took place between February and April 2016 and received 282 responses. Whilst not a novel practice like the OPM process, it serves the similar purpose of enabling discourses from a range of stakeholders to be considered before a Bill's drafting. Doing so supports attempts to ensure the legitimacy of the resulting legislative changes by attempting to understand better the broad range of norms, values, and beliefs held by potentially affected actors. This may lead to action in this context being assumed to be appropriate and desirable. Nevertheless, the consultation paper and the consultation process shaped the discourses that arose in response through its choice of language in presenting the current legislative landscape. The consultation paper posited that legislative reform could 'unlock the potential of publicly held databases to improve the lives of citizens through the uses of data,' including the 'increased use of administrative data for research purposes and official statistics.'[31]

The public consultation generated discourses that sought to bolster the legislative outcome's legitimacy by emphasising both experts and the public's role in the outcome. The inclusivity fostered by this practice should better elucidate the values and beliefs of citizens regarding what is appropriate in this context. Doing so should better enable these values and beliefs to influence legislative changes towards what appropriate data processing practices should be permitted.

[30] Cabinet Office (n 8).

[31] Cabinet Office (n 8).

This would imbue the law with a sense of legitimacy by reflecting a socially constructed system of norms, values, beliefs, and definitions that shape assumptions as to what actions are appropriate. Doing so may help those relying upon DEA to justify their data processing as appropriate and proper. Several experts affected by the current rules on data reuse in research provided such support. The Economic and Social Research Council (ESRC) stated

> *valuable scientific research with potential wider benefit to society undertaken by the UK academic research communities is often delayed, at times cut short, or never attempted in the first place, as a result of the complex legal landscape in data sharing among public authorities.*[32]

Another response echoed this sentiment, stating that

> *access to more varied and better quality data can provide an improved evidence base for researchers…leading to better informed research and more accurate statistics for a range of public benefits…a better understanding of how people live their lives, the patterns of need and use of different services, and the resultant outcomes would support the design and delivery of better services.*[33]

This practice fostered discourses that shaped the key terms and constructs employed in the resulting legislation. There is a significant overlap between the construction of the 'public interest' offered by local authorities and public authorities and the UK government's approach. Lancashire County Council constructed the public benefit as including 'bringing efficiencies for public sector

[32] Cabinet Office, 'Better Use of Data in Government Consultation Responses: Section 6' (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government>, 122.

[33] Cabinet Office, 'Better Use of Data in Government Consultation Responses: Section 8' (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government>, 1.

organisations therefore reducing cost.'[34] The response advocated for creating a framework 'that could be sense checked against a selection of local authority strategies and key aims and also the same for a selection of central government departments.'[35]

Buckinghamshire County Council's criteria for research to be in the public interest included a need for a 'defined and measurable outcome.'[36] However, not all entities agreed with their construction of the 'public interest.' NO2ID argued that 'the point of research is to expand knowledge – a benefit identified by the researcher.' To allow officials to pre-judge purported 'public benefits' would likely bias research towards the theories popular among officials.[37] The Media, Communication and Cultural Studies Association (MeCCSA) echoed this point by arguing that advances in research

> *can only be impeded by prior selection of proposed research by holders of relevant data or their proxies. No accrediting body of the kind proposed can possibly have a credible and legitimate role in pre-judging the value of future research.*[38]

The power asymmetries between these actors led to the dominance of the discourse that better suits the government's aims, whilst still enabling that claim

[34] Cabinet Office, 'Better Use of Data in Government Consultation Responses: Section 2' (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government>, 91.

[35] Ibid, 97.

[36] Ibid, 114.

[37] Cabinet Office (n 34), 206.

[38] Cabinet Office (n 32), 26.

the government 'worked with the research community to produce an effective legal solution to meet [their research] need.'[39]

It is not just clashing discourses that lead to silencing views that contradicted the government's aims in this context. There are examples of views just being ignored. The UCL CLOSER Institute referenced barriers such as 'resource (including insufficient staff with the necessary data management skills and high staff turnover) and out of date data management structures or procedures.' [40] These problems remained unaddressed by the subsequent proposals. Therefore, whilst the presence of experts may legitimise subsequent outcomes, they are not the only 'steer' on the direction of the legislative response. Furthermore, such actors hold little power over the subsequent practices that shaped the DEA. This point is reinforced by a governmental document that set out the need for legislative reform and the potential opportunities this reform could engender. The following section will analyse the discourses generated by the Impact Assessment.

4.3.3 The Impact Assessment

The Impact Assessment,[41] published prior to the legislative process to justify the direction taken by the UK government, strongly emphasises the potential in embracing new technologies. The discourse in the document promotes the

[39] Department for Culture Media & Sport (n 6).

[40] Cabinet Office (n 32), 16.

[41] Department for Culture Media & Sport, 'Digital Economy Bill - Overarching Impact Assessment' (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/573828/2016-11-25_Digital_Economy_Bill_-_Overarching_IA.pdf>.

objective of making Britain 'the most digital nation in the world…a place where technology transforms the economy, society and government.'[42] The revolutionary potential of advancements in technology is prominent in the text. The dominant discourse justifies the appropriate data uses by focussing on the transformative, yet currently inhibited, potential of technological advances. This promotes data reuse as an activity to be encouraged rather than tolerated. The enthusiasm within this discourse fosters a legislative response that is more permissive of data uses than those espousing a more cautious approach. The Impact Assessment outlines its construction of the 'problem under consideration' to justify the necessity of government intervention. Whilst there is no explicit reference to secondary uses of administrative data, they form part of the broader aim of 'supporting the digital transformation of government.'

The framing within the Impact Assessment focuses on the economic benefits, arguing the new legislation is required to make 'sure Britain remains at the forefront of the global 21st-century economy so that our businesses continue to create jobs.'[43] The potential risks to individuals and the uncertainty of research outputs are unacknowledged. Instead, new legislation is portrayed as providing a 'more complete and accurate evidence base to inform analysis and enable better policy design and delivery.'[44] But how far did these discourses continue to dominate the creation of the DEA? The next section investigates how the

[42] Ibid, 1.

[43] Ibid.

[44] Ibid, 2.

legislative process shaped key terms and understandings to promote a particular legislative solution.

4.3.4 The Legislative Process

One of the practices that shape legislation in the UK is the examination of proposed Bills by the Public Bill Committees. Committees can receive written and oral evidence on the proposed approach of Bills and submit and debate proposed amendments. The public and experts may provide evidence to Committee members. The Committee Chair, drawn from the majority party in government, chooses which amendments should be debated. The composition of a Public Bill Committee reflects the relative composition of the political parties within Parliament. Therefore, Members of Parliament (MPs) from the majority party can usually override 'unfriendly' amendments proposed by members of other political parties. This practice legitimises the legislation by promoting the idea that expert interventions improved the scrutineering and amending practices. Doing so enables a further set of actors to try and ensure their values and beliefs are considered during the legislative process, so those who use the law are more likely to be perceived as acting appropriately. There is a melding of democratic practices, where MPs, as representatives of the public, become more knowledgeable and more capable of scrutinising and the input of expert evidence. This promotes a discourse that legislative outcomes are rigorously analysed and less vulnerable to party politics.

The discourse promoted during this process aligned and amplified views that portrayed the existing legal landscape as a barrier that justified legislative intervention. This was represented to require fast action, especially as the barrier

impacted upon the ability to help the public receive the public services they expected. Upon hearing criticisms about the approach of the Bill, one MP stated that 'it is sad that the campaign once again from your organisation is that the perfect must be the enemy of the good.'[45] They further criticised another expert's view as 'disturbing because you are holding up for a perfect world…around privacy and data protection.'[46] The MP then asked

> *at what point do we actually accept that what is proposed in this Bill is actually a good step forward. While it may not be perfect, it is a massive step-change improvement on what we have today.*[47]

This demonstrates the pressure generated by these discourses to do something, to act fast, and accept that the cost is imperfect legislation. It prioritises the discussion of the core objectives, here the capitalisation of advancements and overcoming perceived barriers to reusing data. This prioritisation is at the expense of thoroughly discussing other interests affected by the proposed changes, such as individuals' privacy interests and human rights. The discursive effects thus a legislative reform that can serve these aims rather than bolster the rights-based protections afforded to individuals.

A range of expert stakeholders contributed to the prominence of a discourse that current legislative approaches were causing the UK to fall behind other countries. Hetan Shah, the Executive Director of the Royal Statistical Society, stated that, due to the legislative landscape, the UK was

[45] Digital Economy Bill Deb 11 October 2016, vol 615 col 64.

[46] Digital Economy Bill Deb 11 October 2016, vol 615 col 64.

[47] Digital Economy Bill Deb 11 October 2016, vol 615 col 64.

> *really in danger of being left behind, compared with where other countries are on this agenda…if you look at bodies like New Zealand, Finland and Canada, they all have this ability to access, so we have got to have it.'*[48]

The Economic and Social Research Council, which referred to a report published by the Administrative Data Taskforce in 2012 in its evidence, reinforces this problematisation. It identified the 'vast opportunities that UK's administrative data can provide for research and policy' as a 'rich and largely untapped source of information.'[49] This problematisation portrays the UK as lagging behind other countries that have used their data to 'facilitate important research in an efficient manner.'[50] Such discourses suggest the current legal approach is inefficient and results in the UK missing out on vital opportunities. This problematises the UK as the 'odd one out,' which legitimises a prompt response. Furthermore, it minimises the persuasiveness of claims that certain political ideologies drive the reform. This reduces the scope for opposition by insinuating opposition would be rooted in a desire to see the UK fall behind. The references to the potential detrimental impacts to UK citizens, who would miss out on the benefits that other nations could give their citizens, highlights the apparent importance of the citizen in this legislation reinforces this view. This problematisation justifies a legislative approach that focuses on the benefits for the government and wider society. These discourses are interesting considering arguments that the expertise present in decision making may not be neutral and

[48] Digital Economy Bill Deb 11 October 2016, vol 615 col 82.

[49] Digital Economy Bill Deb 20 October 2016, vol 615 col 246W.

[50] Ibid.

unbiased but lean towards the political.[51] The references to the benefits for policies as well as research aligns with economic rather than scientific claims.

The practice also emphasised the power accorded to the majority political party, who comprised the majority of the Public Bill Committee. This minimising of the views of individuals was criticised whilst the Public Bill Committee was scrutinising the Bill. It was lamented that there was 'little mention of security of privacy, or how such data sharing will comply with obligations around informed consent and the ability to revoke consent.'[52] Nevertheless, the dominance of other discourses meant issues around the role of individuals remained unexplored. During the initial introduction of the Digital Economy Bill, the Secretary of State for Culture, Media, and Sport emphasised the economic benefits of the digital economy and how the 'rate of job creation in digital industries is nearly three times as fast as in the rest of the economy.' The Secretary of State referenced the Manifesto, stating that

> *supporting the digital economy was core to our Manifesto, and that is why the Bill is a central plank of the Government's legislative programme…The Government are working tirelessly to help people and businesses to benefit from digital.*[53]

During its Second Reading, others echoed similar sentiments to reinforce the strength of this discourse. An opposition MP, Chi Onwurah stated the UK is 'Europe's leading digital economy…Now is definitely the time to bring forward a

[51] Alex Faulkner and Lonneke Poort, 'Stretching and Challenging the Boundaries of Law: Varieties of Knowledge in Biotechnologies Regulation' (2017) 55 Minerva 209.

[52] Digital Economy Bill Deb 25 October 2016, col 308.

[53] HC Deb 13 September 2016, vol 614, col 771.

Bill to set out the vision and policies to place the UK at the top of the global digital economy.'[54] Whilst she disagreed that the Bill fulfilled this role, the discourse still focussed upon the digitalisation and advancement agenda.

4.4 EFFECTS

This section considers the effects of the discourses that shaped the final legislative text of the DEA to be the 'solution' to respond to the particular problematisation constructed during the legislative process. The section focuses upon Part 5 of the DEA to analyse four effects. The first is the effect of 'silences' that are present within the DEA. These are the discourses deemed unacceptable, the knowledges that remained unacknowledged, and the unquestioned assumptions that nevertheless informed the approach towards the permissible reuses of data in research. The second is the discursive effects that further shaped understandings of what is permissible and impermissible. In this context, this promoted certain kinds of research as justifying greater access to government-held data. The effect of the approach towards safeguards is analysed, and the section concludes by examining the subjectification effects present in the DEA.

4.4.1 Silences that Promote Reuses of Data in Research

This section, drawing upon the insights of Bacchi[55] discussed in the previous chapter, explores the effects generated by the 'silences' present in the creation of the DEA. These discourses and knowledges were minimised and deemed unacceptable by the dominant discourses promoted by more powerful actors. One

[54] HC Deb 13 September 2016, vol 614, col 778.

[55] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

may also consider 'silences' as created through the unquestioned assumptions and presumptions that justify a particular action over another. The minimisation of discourses may impact the legitimacy of the resulting legislation. It may lead to the omission of certain socially constructed norms, values, and beliefs by which citizens may deem the actions of data users as acceptable and desirable.

The discourses surrounding the DEA further promote research as an activity in the public interest by eliding the potential broadening of access to government-held data and the government's use of research outputs to improve the effectiveness and efficiency of public services. There is a presumption of a strong link within the public consultation between overcoming barriers to secondary uses of data and better policy.[56] It posits that a 'better understanding of how people live their lives, the patterns of need and use of different services, and the resultant outcomes would support the design and delivery of better services.' [57] The promotion of this discourse glosses over the further steps that must be taken for research to influence and shape policy despite this being raised during the public consultation process.[58]

In silencing discourses that considered barriers other than the lack of a specific legal gateway, the DEA's approach implies the lack of a specific legal gateway was the only barrier to be overcome. This selective construction of the problem minimises the impact of previous comments where a government

[56] Cabinet Office (n 8).

[57] Ibid, 27.

[58] Cabinet Office (n 32), 15.

representative admitted it 'may well be true'[59] that the greater barrier was a cultural one rather than a legal one. Nevertheless, they argued that the DEA was 'the first step towards changing the culture.'[60] This discourse, promulgated by the Government in the wealth of documents analysed thus far, subjugates other views on the true 'problem' of facilitating secondary data uses. The subjugated views are those proposed by the experts, whose contributions sought to add further legitimacy to the legislative outcome. CLOSER spoke of being hindered by cultural barriers within government departments and 'disparate and opaque data access processes and procedures.'[61] The effect is to create a government-focused solution that retains its veneer of the approval of a wide range of directly affected actors.

A failure to engage with those who criticised the attempts to be more inclusive during the legislative process is a further silence that promotes the government's preferred discourses. Some actors within the Open Policy Making process complained of problems arising from 'decisions or policies being announced at short notice from other parts of government being prioritised.' [62] The government minister in attendance did not acknowledge these comments. Instead, their reflections reiterated the potential benefits of opening up more government-held data to researchers and the 'beneficial' and 'ground-breaking' nature of the

[59] Involve, 'De-Identified Data for Research | 15 May 2014 | Meeting Note' (2014) <https://web.archive.org/web/20181129073618/http://datasharing.org.uk/category/meeting-notes/page/2/index.html>.

[60] Ibid.

[61] Cabinet Office (n 32), 15.

[62] Involve, 'Plenary Workshop with Francis Maude - Meeting Note' (2014) <https://web.archive.org/web/20160513051853/http://datasharing.org.uk/2014/11/07/plenary-workshop-with-francis-maude-22-oct-2014-meeting-note/>.

collaboration fostered by this process.[63] The prioritised discourses and subjugated knowledges demonstrate the power the government and its actors retained over the practice.

The Codes of Practice, which were vital to understanding how the proposed powers would work, were yet to be drafted during the consultation and OPM process practices. The consultation exercise was heavily criticised for other reasons. In his written evidence, Dr Whitley stated that 'such poorly designed consultation questions may also have affected the quality of submissions.'[64] Moreover, two sets of proposals were added to the consultation at a very late stage, a development Dr Whitley perceived, 'given the spirit of the [OPM] process' as 'unfortunate.'[65] He felt the rushed analysis of the consultation generated 'concerns' about their use in the scrutiny process, where certain MPs used the number of respondents to suggest a 'broad public support' for the sharing of data proposed under the Bill.[66]

The practices and discourses analysed in this section demonstrate how the government's actions during the public consultation undermined that practice's ability to inform the direction of any potential legislative reform productively. At all stages of the consultation process, weaknesses were identified that undermined

[63] Ibid.

[64] Edgar A. Whitley, *Written evidence (LEG0023)*, in House of Lords Constitution Committee, 'The Legislative Process: Preparing Legislation for Parliament Oral and Written Evidence' (2016) <https://www.parliament.uk/globalassets/documents/lords-committees/constitution/legislative-process-inquiry/evidence-volume-legislative-process-stage-1-word.pdf>.

[65] Ibid.

[66] Ibid.

the ability of external actors to feed their knowledge into the legislative process. Before the release of the consultation process, the External Advisory Group to the consultation process was highly critical of many aspects of the process.[67] They felt that insufficient care was taken to ensure the document's layout could appropriately gather the type of feedback and input supposedly wanted by the government. During the consultation process, insufficient time was given to appropriately analyse the responses, particularly considering the document's formatting issues. Finally, by introducing proposals at a very late stage, the government further undermined the ability of external stakeholders to consider the options and respond with their considered views. Nevertheless, these shortcomings were not recognised in the government's response to the consultation.

Some of the practices identified in this chapter minimise the 'human' element inherent within administrative data. For example, certain evidence and discourses focus upon the 'vast opportunities that UK's administrative data can provide for research and policy'[68] as a 'rich and largely untapped source of information.'[69] This positive view of the potential for secondary data uses fails to acknowledge a critical aspect; the public's views on using their data and their response to these changes. Doing so creates a risk that the resulting legislation

[67] External Advisory Group, 'Report from the External Advisory Group for the Better Use of Data in Government Consultation' (2016) <https://web.archive.org/web/20180901171540/http://datasharing.org.uk/2016/07/15/read-the-external-advisory-group-report/index.html>.

[68] Digital Economy Bill Deb 20 October 2016, vol 615, col 246W.

[69] Ibid.

may lack legitimacy by failing to consider the norms and values used by citizens to determine if an entity's actions are deemed appropriate. Nevertheless, those who justify their actions as lawful may have their actions deemed inappropriate or undesirable by citizens through the latter's reference to a socially constructed set of norms, values, and beliefs that was not fully considered during the legislative process. Within the public consultation, Mydex's response noted that

> *there is nothing in the legislative proposal to genuinely make citizens digitally enabled or able to act independently, from which self-reliance and innovation will come, as well as a more secure, more trustworthy society of self-sovereign individuals.*[70]

The minimising of the individuals' views was criticised whilst the Public Bill Committee scrutinised the Bill.[71] The extent to which individuals are considered may be perceived as one of the 'silences' left unproblematic by the dominant discourses that constructed the DEA's approach. This is despite the heavy reliance on the 'public interest' in determining permissible reuses of data. The inclusion of the public in such decisions, to help elucidate the 'public interest' in this context, was raised during the public consultation. Hubs for Trials Methodology Research Health Informatics Group asked

> *Why not include members of the public and members of the research community in this assessment [of what is in the public interest]? The involvement of patients and public during the development and conduct of research projects, ethical review, funding decisions,*

[70] Cabinet Office (n 34), 80.

[71] Digital Economy Bill Deb 11 October 2016, vol 615 col 73 and Digital Economy Bill Deb 27 October 2016, vol 616 col 338.

> *dissemination and implementation, has become routine practice in medical research.'*[72]

What remains 'unsaid' is the risk that, without further detail or guidance, the needs of those approving research may be used as a proxy for the 'views of the public.' The NO2ID consultation response elucidated the potential benefits from more inclusive actions as, in their view, 'giving control of personal data back to citizens will improve the quality and efficiency of public services.'[73] However, there was no engagement with this point. This is even more critical in this context where Citizens Advice's response recognised during the consultation that

> *[A]s any government data sharing initiative will inevitably start from a low base of trust, we feel that the measures to reassure trust and accountability should be strengthened to take account of citizens' concerns and requirements.*[74]

The silenced discourses were not the only influence on the DEA. The next section will consider the discursive effects that arise from the DEA, shaping what is deemed acceptable and unacceptable in the context of reusing administrative data in research.

4.4.2 Discursive Effects Construct Research as an Activity that is Permissible Where it Suits Governmental Aims

One effect of the legislative process was to construct the key term of 'research' to enable certain types of research to be deemed 'worthy' of greater access to

[72] Cabinet Office, 'Better Use of Data in Government Consultation Responses: Section 3' (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government> accessed April 29, 2021, 287.

[73] Cabinet Office (n 34), 212.

[74] Cabinet Office (n 72), 181.

administrative data. This effect is present in the decision to create a permissive gateway.

S64(1) DEA merely creates a permissive gateway where 'information held by a public authority in connection with the authority's function may be disclosed to another person for the purpose of research which is being or is to be carried out.' The effect of this places significant power in the hands of the public authorities that currently hold the data as, according to the Open Data Institute, 'access is still dependent on the wishes of the public body maintaining the data asset.'[75] The discourse promoted by the government was that the 'real' problem was confusion as to what powers were available, thus the permissive gateway response was sufficient to 'help unblock the flow of data.'[76] This dominates opposing discourses that felt proposals should go further than the permissive power as currently drafted.[77] This 'truth' of the dominating discourse is supported by a 2016 document issued by the government that claimed a 'permissive power across all public authorities…to share/link de-identified data explicitly for the purposes of research in the public interest' was a 'key' area of agreement from the 2014-15 OPM process.[78]

The discursive effect facilitates the governmental assumptions concerning what research is worthy of greater access to administrative data. The needs and

[75] Cabinet Office (n 72), 240.

[76] Department for Culture Media & Sport (n 6).

[77] Cabinet Office, 'Better Use of Data in Government - Consultation: A Government Summary of Responses' (2016), para 79.

[78] Cabinet Office, 'Data Sharing Proposals Paper' (2016) <https://docs.google.com/document/d/1d9aoCe9e76yMtNyfkv2QjEAQOh6dyuCYdCeWb4BXFNY/edit>.

wants of the UK government still shape the understanding of the 'opportunities' made possible by legislative reform. The government's response to the consultation acknowledged that

> *a number of responses, particularly from the research community, expressed the view that the proposals should go further than the permissive power as currently drafted [and] there should be greater incentives for public authorities to share data.*[79]

Nevertheless, the power wielded by governmental actors meant no changes were made.

The DEA's permissive gateway assumes the worthiness of research is best determined by public authorities. The prominent discourses propose there is a 'presumption to share.'[80] The discursive effect implies that refusals to share data result from the unworthiness of the proposed research, rather than any bias or caution from the public authorities. Underlying this legislative approach is the insinuation that public authorities should have the 'final say' over what research projects should be authorised. Doing so enables them to consider reasons above and beyond the safeguards present in the DEA. Failing to elucidate what may prevent the approval of research demonstrates that reasons other than the 'safeguards' included in the DEA warrant limitations on research using administrative data. The effect, nevertheless, facilitates the appearance of a piece of legislation, and the lawmaking powers behind it, as supportive of research and

[79] Cabinet Office (n 77), para 79.

[80] Department for Culture Media & Sport (n 6).

encouraging researchers to use this gateway to fully realise the opportunities offered by secondary uses of administrative data.

The accreditation criteria that the researcher, their research project, and the entity sharing data must fulfil further facilitates only certain types of research. This constructs the key term of 'research' as an activity that may benefit from the DEA in a certain way. By requiring research to be 'in the public interest,' a certain type of research is prioritised. The system was criticised by the MeCCSA as 'wholly incompatible with free academic debate.'[81] The proposed accreditation body was further condemned by MeCCSA due 'its capacity improperly to pre-decide who would have access to data and how it would be used.'[82] The effect promotes a construction of research in the public interest that produces more tangible, quantifiable benefits. The criteria do not promote academia in its own right, nor academic debate, but rather research that can best demonstrate its tangible contribution to the goals and aims of the authority holding the data. According to the Big Data Network (BDN), this construction excludes 'curiosity-driven or speculative investigations,' overlooking that '[p]ublic benefit is not necessarily either short-term or subject to the realisation of immediate value.'[83] This effect is despite East Sussex County Council's response that emphasised that

> *…any evaluation should also allow for speculative research that is typically harder to justify in the public interest because the outcome is*

[81] Cabinet Office (n 32), 24.

[82] Ibid.

[83] Ibid, 77.

> *much harder to predict…[it] has the potential to be ground-breaking in the future and this should be recognised.*[84]

BDN's consultation response echoes this sentiment through its preference for a clause,

> *which recognises that data drives economic and social research, that this research has value, and that restrictions to data sharing suppresses research that is of public benefit, regardless of whether those benefits are immediate or to be realised in the fullness of time.*[85]

A question to be explored in the empirical work is how far participants' views reflect these views and beliefs as to 'worthy' research found within the practices shaping the DEA. This may enable insights into the participants' perceived legitimacy of the provisions of the DEA.

Other aspects of the accreditation criteria[86] further support a certain construction of research that justifies this permissive approach to data sharing. This promotes research where it is more likely to create more tangible, concrete benefits. The accreditation conditions include the requirement that the researcher demonstrates evidence of 'suitable research qualifications and/or experience.'[87] This is defined either as an undergraduate degree 'including a significant proportion of mathematics or statistics' or 'at least 3 years quantitative research.'[88]

---

[84] Ibid, 12.

[85] Ibid, 79.

[86] UK Statistics Authority, 'Data Sharing Code of Practice and Accreditation Criteria: Research' (2016) <https://www.gov.uk/government/consultations/digital-economy-act-part-5-data-sharing-codes-and-regulations/research-code-of-practice-and-accreditation-criteria>.

[87] Ibid, para 25.1.

[88] Ibid.

This implies that all appropriate and justifiable research is quantitative, potentially excluding researchers with a qualitative background.

These requirements bias a construction of research in the public interest towards quantitative research. The dominant discourse that supported this position was promulgated more by the public authorities than academic entities. East Sussex County Council's response supported research that could show how its benefits could be monitored and assessed through quantitative and qualitative measures and would oppose the secondary use of administrative data in research were 'solely or primarily for the achievement of qualifications.'[89] Furthermore, the definition of 'public interest' is significantly influenced by the definition of 'public good' set out in the Statistics and Registration Service Act 2007. The presumption underlying this reliance on the Statistics and Registration Service Act is that the two instruments cover similar research types to allow this common understanding of the public interest. Therefore, the kind of research facilitated through the DEA will be the type of research also governed by the statistics and quantitative research governed by the Statistics and Registration Service Act, at the potential expense of more exploratory and more qualitative-based research.

The discursive effect of the term 'public interest' promotes the assumption that a singular 'public interest' in this context exists, despite the heterogeneity of the wants, needs, and desires of those in society. The potential fallacy was raised during the consultation process, where the government was criticised for presuming only positive outcomes would arise from the greater reuse of data in

[89] Cabinet Office (n 32), 3.

research. The discursive effect limits discussion as to whether the public interest is served where data sharing 'results in some people no longer receiving the service, or say 15% of those whose data was shared were deemed not eligible for the service?'[90] Digitaldefendme's response echoes this concern that

> *using an individual's data should not be punitive or harmful to that individual, but in the case of families, what may be good for one could be seen as detrimental to other family members.*[91]

The effect of this range of inclusive practices before the passing of the DEA promotes the legislation as reflecting a more complete and accurate construction of the public interest.

Nevertheless, upon closer examination, the discourses of the public authorities in this area dominate the term's construction. Cheshire East Council[92] also focussed on the possibility of tangible benefits to services they provide and not private sector profit and Buckinghamshire Council's support for research with a 'defined and measurable outcome.'[93] This overshadows other constructions, such as that provided by NO2ID, which advocated for researchers to identify the benefit otherwise it would enable 'officials to pre-judge purported public benefits and would likely bias research towards theories popular among officials.'[94] This fear about other uses of the public interest to promote the kind of research supportive

[90] Cabinet Office, 'Better Use of Data in Government Consultation Responses: Section 5' (2016), 45.

[91] Cabinet Office (n 34), 153.

[92] Ibid, 108.

[93] Ibid, 114.

[94] Ibid, 216.

of governmental, rather than societal, wishes is partially echoed by the National Foundation of Educational Research. Its response stressed that 'sometimes the purpose of undertaking new research is to bring a particular issue to the attention of policymakers and the public.'[95] The dominant discourse operates as a further bar on allowing citizens and actors outside of government from helping to determine what is and is not appropriate in this context. Instead, it concentrates further power in the hands of governmental actors.

Furthermore, by allowing non-academic actors to determine what research should be permissible under the DEA, there is an assumption that such actors have the expertise to determine what research is in the 'public interest.' As a result, just as science informs policy, policy now informs the direction that science should take through the 'public interest' as a criterion for the appropriateness of research and academics.

4.4.3 The Effects Lead to a Construction of Adequate Safeguards That Facilitates Data Reuse

The permissive nature of the gateway created in the DEA is not the only limit on the type of research that may reuse administrative data. This section focuses on the use of 'safeguards' to limit the types of research that are permissible. Certain practices that occurred during the legislation process shaped their presence in the DEA. This section explores the effects generated from the assumptions that underpin the safeguards referenced in the DEA. These safeguards are assumed to ensure that the broadened access to administrative data only facilitates

[95] Cabinet Office, 'Better Use of Data in Government Consultation Responses: Section 7' (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government>, 132.

'appropriate and ethical' data science techniques to 'help tailor interventions to the public.'[96]

Under s64 DEA, information may not be disclosed unless six conditions are met. The Explanatory Notes clarify that these conditions are for 'safeguarding the privacy of individuals.'[97] There is a presumption here that, where the aim is to facilitate appropriate and ethical secondary uses of administrative data in research, only individuals' privacy interests must be safeguarded. This is despite a wealth of academic literature exploring how modern data processing practices impact group privacy interests.[98] The effect suggests other concerns are not relevant to determining what types of research should be permitted, increasing the scope of projects that may be approved.

The DEA only permits the sharing of data where a person's identity is not specified in the information and 'it is not reasonably likely that the person's identity will be deduced from the information (whether by itself or taken together with other information.)'[99] The Explanatory Notes clarifies that 'both non-personal and personal information can be disclosed pursuant to section 64(1)'[100] and so anonymisation is not required. The effect implies a consensus as to what protections reach the threshold of 'reasonably likely,' which allows a certain level of risk to be inherent in the reuse of data in research. Reasonableness also

[96] Cabinet Office (n 8).

[97] Explanatory Notes to the Digital Economy Act 2017.

[98] Brent Mittelstadt, 'From Individual to Group Privacy in Big Data Analytics' (2017) 30 Philosophy & Technology 475.

[99] S64(3) Digital Economy Act 2017.

[100] Explanatory Notes to the Digital Economy Act 2017.

moderates the requirements regarding minimising the risk of accidental disclosure, re-identification, and deliberate unlawful disclosure. The effect of this language choice aligns societally acceptable secondary data uses with the stringency of this requirement. It assumes that a data discloser can know enough about the potential risks involved to make an informed decision about the 'reasonable steps' required in each context to minimise the risk of further disclosure. This is despite the inherent unknowns associated with data processing, particularly where large data sets are used, or datasets are combined. Moreover, the focus on preventing disclosures of an individual's information is based on the presumption that only when an individual's identity is disclosed can the secondary use of data be deemed unacceptable and unethical and thus justify a legislative response.

The final safeguard listed in s64 DEA is the need for those involved in this secondary use of data to 'have regard' to the Code of Practice whilst disclosing data.[101] This presumes that 'have regard' is sufficiently stringent to ensure that only acceptable and ethical uses of data are allowed, thus safeguarding individuals' privacy interests. The interactions between the code and the law are assumed not to engender a level of complexity that would undermine the justifications for the DEA as resolving an uncertain and complex legislative landscape. This is despite warnings by the Information Commissioner in her

[101] S64(10) Digital Economy Act 2017.

written evidence that the interactions between different pieces of legislation risked creating further confusion.[102]

The assumptions underpinning the choice of safeguards facilitate 'acceptable and ethical uses of data.' This allows a level of risk that actors involved in the lawmaking process presume to be reasonable. The effect is to also minimise the relevant harms to an individual's privacy interests, despite a wealth of academic literature emphasising the range of harms such data processing may cause[103] and the nature of the risks individuals may face because of this secondary use of data.[104]

4.4.4 Subjectification Effects

The subjectification effects refer to how certain groups or individuals are constructed a certain way because of the problematisation. The discourses that shaped the DEA as the solution to overcoming unjustified barriers to secondary data uses foster certain subjectification effects. These effects concern how citizens are constructed as a result of the law. In this context, the subjectification effects construct an understanding of a citizen that promotes harmony between their views, desires, and wishes with the DEA's approach. This section explores the dominant discourses that shaped individuals as supportive yet passive actors

[102] Digital Economy Bill Deb 13 October 2016, vol 615 col 116W.

[103] Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences* (SAGE Publications, 2017).

[104] Mittelstadt (n 98).

within the process. This produced legislation that minimises the active involvement of individuals.

One of the key justifications utilised by the government for this reform was the public's expectations to employ advances to deliver more effective and efficient services. This aligns with governmental intentions regarding data reuses yet fails to engage with many responses to the public consultation that questioned the lack of a role for individuals. The ADRN noted that the public might perceive data collected by public authorities as a public resource that should be more widely used to maximise its utility. Nevertheless, they equally expect 'robust and proportionate safeguards, in recognition of the public interests served by protecting individual privacy.'[105] Constructing citizens as supportive enables the government to moderate the proportionality of these safeguards to potentially allow for more secondary uses of administrative data in research. It minimises discourses that attempt to envisage more 'active citizen being more responsible and engaged' [106] to pursue the true objective as developed through the problematisation identified in this chapter; the removal of barriers to allow for the full capitalisation of technological advancements related to the secondary uses of data.

There are also subjectification effects that construct a certain kind of researcher. This is evident in the type of research promoted in the practices that shaped the DEA that aligns with the idea of research in the 'public interest' being

[105] Cabinet Office (n 32), 112.

[106] Cabinet Office (n 32), 137.

the kind of research that produces tangible and measurable outcomes. There is an assumption of a strong alignment between the objectives of researchers and public authorities in this context towards such outcomes, minimising the scope for more curiosity-driven research to qualify as in 'the public interest' and benefit from access to administrative data.

4.5 Concluding Remarks

Within this chapter, the analysis of the problematisation has shed light on the DEA being crafted as a solution to unjustified barriers to secondary data uses. The practices that shaped the problematisation have sought to emphasise the opportunities and benefits presented by secondary uses of data and the downsides to the UK falling behind the progress made by other nations to justify the legislative response. However, the discourses arising from the practices, such as the OPM process, the public consultation, and the Parliamentary scrutiny of the Bill, were not always harmonious. Practices that included a broad range of actors, especially from outside government, demonstrated differences in opinion in the underlying problem and the most appropriate means to solve it. Nevertheless, the differences in the power wielded by actors that took part in the process help explain why the approach of the DEA so closely aligns with the problematisation regarding barriers and missing out on opportunities that other nations are capitalising upon to benefit their citizens. The power wielded by governmental actors throughout this process undermined the influence of the inclusive practices, which introduced other problems and ways of thinking into the discourses shaping the DEA, over the legislative response.

The analysis suggests that discourses within these practices, which aligned with the government's pre-determined objectives and aims, were taken forward in a way that enabled the government to claim it had listened to a wide range of actors, both experts and from the public. In doing so, it served to legitimise its response as democratic and even-handed. This imbues the approach of the DEA, particularly where it appears to draw upon societal values to determine appropriate secondary uses of data, with further legitimacy and alignment with the desires of citizens. This is evident in the safeguards requiring research to be in the public interest or accredited to ensure it is appropriate. These appear to protect the interests of individuals, but this chapter has sought to expose how these 'safeguards' may also operate to shape appropriate secondary uses of data to promote certain types of research for certain purposes that fit governmental needs, rather than societal acceptability. However, it is unclear how far this focus on opportunities and benefits fostered by greater secondary uses of data fails to align with societal expectations. The next part of this thesis focuses upon societal expectations towards secondary uses of administrative data in social science research through empirical research.

Part II: The Empirical Findings

5 AN INTERLUDE: CONNECTING THE DOCTRINAL AND THE EMPIRICAL WORK

5.1 INTRODUCTION

This chapter serves as a 'bridge' between the doctrinal analysis and the empirical work. Firstly, it considers the findings from Part I, which suggests societal expectations are deemed important in constructing the legal approach but are dominated by subsequent discourses proposed by powerful actors during the legislative process. This potentially has implications for the law's legitimacy as it indicates it may not be deemed an appropriate way to govern secondary uses of administrative data in research by reference to a society's system of norms, beliefs, and opinions. By drawing upon the works of Poort,[1] Nonet and Selznick,[2] and Fuller,[3] this chapter explores the negative implications for the quality and legitimacy of the law produced. This assessment requires a further investigation of societal expectations to shed light upon how far they seem to diverge from the legal approach analysed in Part I. However, it is not only for normative reasons that societal expectations remain salient after legislation is passed. Examples of legally permissible data reuses that have triggered a public backlash due to their misalignment with societal expectations demonstrate the practical implications of

[1] Lonneke Poort, 'Consensus & Controversies: An Interactive Legislative Approach to Animal Biotechnology in Denmark, Switzerland and the Netherlands' (PhD thesis Tilburg University, 2011).

[2] Philippe Nonet and Philip Selznick, *Law and Society in Transition: Toward Responsive Law* (Transaction Publishers 2017).

[3] Lon L Fuller and Kenneth I Winston, *The Principles of Social Order: Selected Essays of Lon L. Fuller* (Hart Pub 2001).

divergences between law and societal views.[4] Public backlashes trigger mistrust, which potentially jeopardises the ability to capitalise on the opportunities presented by innovative data processing practices and undermine the law's quality.

5.2 BUILDING BRIDGES BETWEEN THE LAW AND SOCIETAL EXPECTATIONS: THE DOCTRINAL UNDERPINNINGS

In attributing value and prominence to the views and expectations of citizens within society, the thesis draws upon theories that consider how societal attitudes may guide our assessment as to the 'quality' of a given law. Where laws fail to fit with societal expectations, this may undermine their quality and legitimacy. Nonet and Selznick note that

> *characteristically, legal theories are built upon implicit theories of authority…[and] the underlying mood – a disaffection from what has seemed for so long and to so many a crowning achievement of Western civilisation – cannot be dismissed.*[5]

By introducing a social science perspective to jurisprudential conceptions, they argue that it is preferable to explore to what extent and under what conditions connections between law and the state, law and rules, and law and moral aspiration occur. This is rooted in the idea that a 'legal order has many dimensions' and that 'univalent conceptions of law as "an ordinance of reason for the common good"' is to 'recreate the parable of the blind man and the elephant.'[6] This comparison highlights how people are biased towards claiming to know an

[4] Fiona Godlee, 'What Can We Salvage from Care.data?' (2016) 354 BMJ 3907.

[5] Nonet and Selznick (n 2), 4.

[6] Nonet and Selznick (n 2), 9.

absolute truth based upon their limited, subjective experiences and ignore others limited and subjective experiences, which may also be true. Poort[7] proposes three dimensions of 'fit' as a framework for assessing the quality of law: the 'fit with law'; 'fit with context'; and the ability of the law to cope with problems. Firstly, the 'fit with law' relates to the relationship between the law analysed and the wider legal order. Secondly, the 'fit with context' dimension is highly relevant to this thesis as it recognises that the quality of law depends on its fit with the social context. The social context is defined in terms of societal input in developing norms. Without such input, the legal norms may struggle to adjust to societal demands and societal problems.[8]

Norms provide reasons to act in particular ways, driven by underlying values rooted in society's culture.[9] Selznick argues that 'for a well-ordered legal system, nothing is more important than social support.'[10] Poort draws upon this to discuss the third dimension of the law's ability to 'cope with problems.' This is relevant to this thesis, which concerns the governance of a highly complex and rapidly evolving set of practices. Here, innovative technologies make determining how far a law may 'fit' with a given social context increasingly difficult. In response, rather than to construct law as highly prescriptive, this thesis draws

[7] Poort (n 1).

[8] Nonet and Selznick (n 2), 16.

[9] Reza Banakar, 'Can Legal Sociology Account for the Normativity of Law?' in Matthias Baier (ed), *Social and Legal Norms: Towards a Socio-Legal Understanding of Normativity* (Routledge 2016).

[10] Philip Selznick, "Law in Context' Revisited' (2003) 30 Journal of Law and Society 177, 178.

upon the theory of 'responsive law' articulated by Nonet and Selznick.[11] This theory enables the law to become 'an instrument for change…open to reconstruction in the light of how those who are governed perceive their rights and reassess their moral commitments.'[12]' This type of law 'perceives social pressures as sources of knowledge and opportunities for self-correction.'[13] This resonates with the literature on social licences, which can help sustain and legitimise particular practices where there are differences between what is legally permissible and societally acceptable.[14] These enable decision-makers to 'interpret and reformulate rules in light of their actual consequences [considering] broader principles of law, justice, and public policy.'[15] Therefore, the law becomes a 'facilitator of response to social needs and aspirations'[16] to 'guide and legitimate political and legal change.'[17]

5.2.1 Creating Meaningful Space for Public Participation

The previous section's construction of the law suggests that legal language becomes important due to the legislative provisions' intent to 'shape the world of

[11] Nonet and Selznick (n 2).

[12] Ibid.

[13] Ibid,77.

[14] Mhairi Aitken and others, 'Establishing a Social Licence for Financial Technology: Reflections on the Role of the Private Sector in Pursuing Ethical Data Practices' (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720908892>.

[15] Nonet and Selznick (n 2),16.

[16] Ibid,18.

[17] Ibid, XVII.

the present and in particular the future.'[18] Law moves from operating as an instrument for issuing commands by the legislature to operating as a symbol to facilitate ongoing interactive processes to create and implement legislation.[19] This allows the law to remain open to interpretation and 'expressive of fundamental values.'[20] This form of law and lawmaking implies participation in legal interpretation to emphasise the state's active role and responsibility to facilitate public participation.[21] Selznick rejects the idea of abandoning the distinction between positive law and its social contexts. Instead he emphasises that positive law is nevertheless continuous with other norms and practices that affect the character of law.[22] Such ideas emphasise the continued importance of exploring and incorporating societal views and expectations through key terms and norms utilised within the law.

The importance of societal values in shaping the law aligns with the argument of Nonet and Selznick that 'a distinctive feature of responsive law is the search for implicit values in rules and policies.'[23] The articulation of these values thus offers authoritative criteria for criticising existing rules and generating new

---

[18] Reuben Binns, 'Analogies and Disanalogies between Machine-Driven and Human-Driven Legal Judgement' (2021) 1 Journal of Cross-disciplinary Research in Computational Law <https://journalcrcl.org/crcl/article/view/5>.

[19] Bart van Klink, 'Legislation, Communication, and Authority. How to Account for the Bindingness of Law?' in A Daniel Oliver-Lalana (ed.), *Conceptions and Misconceptions of Legislation* (Springer International Publishing 2019).

[20] Ibid.

[21] Peter Vincent-Jones, 'Responsive Law and Governance in Public Services Provision: A Future for the Local Contracting State' (2003) 61 The Modern Law Review 362.

[22] Selznick (n 10), 178.

[23] Nonet and Selznick (n 2),79.

ones. This demonstrates the multiple roles that values have within this context. These values are held by individuals, and matter to them, and can be used to unify practices into normative systems, such as law.[24] In doing so, one sees the centrality of purpose in helping sustain the legitimacy of law as a means of preventing officials from retreating behind rules to evade responsibility. Instead, where accountability is to more general ends, this is insufficient to insulate them from criticism. The vagueness potentially fostered by the abstract level at which the law's purpose exists is countered through 'missions.' Missions translate the general purpose into specific objectives.[25] This allows the law's general purpose to 'promote the public interest' to be refined in specific contexts, responding to difficulties arising from using abstract ideas to provide guidance and evaluation standards.

These ideas also align with the framework proposed by Bacchi,[26] whose approach was so influential to the analysis within Part I of this thesis. As part of understanding the problematisation at play, she advocates a consideration of where the representation is produced, disseminated, and defended and how it could be questioned, disrupted, or replaced. After passing a legislative solution, the representation remains within the spheres of those who govern data, potentially excluding data subjects from questioning these uses. The context of reusing administrative data in social science research exacerbates this concern.

[24] Sanne Taekema, 'Two Realist Idealists Theorizing the Rule of Law' (2019) 11 Hague Journal on the Rule of Law 301.

[25] Nonet and Selznick (n 2), 83.

[26] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

This data is often passively collected about individuals, and there is no legal requirement to inform a data subject of reuses of their data in research. Therefore, this heightens the risk that the law, as interpreted and used by data controllers, will diverge further from societal expectations due to the difficulties in questioning the use of the law to facilitate reuses of data in research.

5.2.2 The Importance of Legitimacy

Poort and the theory of responsive law find that law must demonstrate legitimacy to be of good quality. Legitimacy is relevant to Nonet and Selznick's articulation of responsive law. A rule may meet the 'pedigree test' of validity through bearing the 'stamp of official authority,' but it is still open to reassessment 'in the light of its consequences for the values at stake.'[27] However, they associate legitimacy as the central concern of autonomous law, with a definition of legitimacy that relates to the certification of the authority of rules and decisions, rather than ensuring that institutions have the competence to fulfil their mandates.[28] This thesis draws upon a different understanding of legitimacy, responding to some of the implied criticisms.

The thesis defines legitimacy as a concept based upon a relationship between an organisation and its environment whereby there is a 'generalised perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and

[27] Nonet and Selznick (n 2), 82.

[28] Nonet and Selznick (n 2), 104.

definitions.'[29] This helps to justify the exercise of political power where it aligns with the principles and ideals acceptable to citizens as reasonable and rational.[30] Within that definition, there is some recognition of the need to consider the competence of institutions to make decisions that fulfil the law's purpose and provide substantive justice to individuals. It also chimes with Nonet and Selznick's understanding of purposive law and regulation, using an inclusive conception of the legal process. Consequently, the law becomes a 'problem-solving, facilitative enterprise.'[31] In drawing upon socially constructed norms, Poort[32] links the quality of legal norms to substantive justifications and popular legitimacy. The latter is based upon Fuller's[33] work on the requirements of a legitimate legal system. A legal system cannot function solely through vertical relationships that create formal rules of duty. It must also consider the 'shared commitments' of a society regarding its aspirations and convictions.

This aligns with van Klink's arguments that legislation expresses values and offers a vocabulary that affects how people perceive reality.[34] Law earns its authority by including a range of actors and being responsive to needs. This presupposes a relationship of reciprocity between the legislature and citizens and

[29] Mark Suchman, 'Managing Legitimacy: Strategic and Institutional Approaches' (1995) 20 Academy of Management 571, 574.

[30] Mark Taylor and Tess Whitton, 'Public Interest, Health Research and Data Protection Law: Establishing a Legitimate Trade-off between Individual Control and Research Access to Health Data' (2020) 9 Laws 6.

[31] Nonet and Selznick (n 2), 110.

[32] Poort (n 1), 23.

[33] Fuller and Winston (n 3).

[34] van Klink (n 19).

its compliance with the principles of lawmaking.[35] Therefore, the value of societal expectations and views lies in the potential for enhancing the quality and legitimacy of law. In doing so, the legal arena takes on a political dimension where participation permits more than the vindication of individual rights based on recognised rules. [36] Nonet and Selznick argue that this enlargement of participation increases the democratic worth of law and contributes to the competence of legal institutions.[37] It is acknowledged from the start that there is a delicate balance to be struck to ward against the risk of shifts resulting from 'whims, ignorance, and manipulation.'[38] Part III of the thesis details this further by exploring potential responses to the divergences identified in this thesis and their potential causes.

5.2.3 Why Societal Expectations Matter

Innovations, such as those that triggered the legal reforms discussed in Part I, rely on societal acceptance, yet their inherent uncertainties trigger diverse value-laden perspectives. Therefore, many argue in favour of public engagement before legal reform, particularly in response to technological changes. The underlying aim is to increase the legitimacy and accountability of the outcomes, empower participants, and develop more effective responses. Both the General Data Protection Regulation (GDPR) and the Digital Economy Act 2017 (DEA) utilised public engagement mechanisms, through consultations and the Open Policy

[35] Ibid.

[36] Nonet and Selznick (n 2), 96.

[37] Nonet and Selznick (n 2), 98.

[38] Henry Tam, 'Government with the People' in Henry Tam (ed), *Whose Government Is It?: The Renewal of State-Citizen Cooperation* (Bristol University Press 2020).

Making process, to foster the citizens' inclusion. The practices adopted by the actors involved in the legislative process supports the idea that societal views and societal acceptance were an important aspect of ensuring the quality of any legislative outcomes.

Both the GDPR and the DEA used consultations before the drafting of the legislation. This enabled the views of citizens and other interested actors to be sought. Furthermore, the Open Policy Making process, undertaken before the introduction of the Digital Economy Bill, aimed to incorporate better a range of views into the legislative process, better securing societal acceptance. Furthermore, the idea that those affected by political decisions should influence decision-making underpins democratic societies. Both the Members of the European Parliament involved in creating the GDPR, and the Members of Parliament involved in enacting the UK's DEA are supposed to represent the electorate's interests. This is evident in the European Parliament's approach during the drafting of the GDPR. Its restrictive approach towards reusing data in research aligned more closely with some of the concerns raised by individuals during the consultation process.[39] The final practice that supports the importance of societal expectations and acceptance are the terms adopted in the legislation that justify the legislative approach by recourse to the public interest and societal expectations.

Whilst the practices identified above support the apparent importance attached to societal views and expectations towards securing the law's legitimacy,

[39] See Chapter 2 of the thesis for further detail of this approach.

the analysis in Part 1 of this thesis demonstrate that the outcomes of public engagement are often at the mercy of other actors and subsequent practices. These actors' abilities to dominate the discourses during the legislative process enable them to utilise societal views to support the discourses that align most with their desires. The dangers of poor public engagement practices are heightened in the context of technological innovation due to its increasing perception as part of the political sphere.[40] The literature concerning social licences highlights the problems that can arise where legal practices are societally unacceptable.[41] Therefore, it becomes necessary to uncover any divergences between the legal approach to the reuse of administrative data in social science research and societal expectations towards this activity. If divergences are identified, further empirical work will be necessary to explore the potential causes. For example, does the law cover all aspects of the data processing activity that individuals view as requiring legal governance? Does the law include definitions of the key terms that people would expect to apply to secondary uses of administrative data in social science research? Are the 'silences' identified Part I critical to individuals' support or opposition? The subsequent investigation of attitudes and expectations can help guide the interpretation of the key terms in the law that justify its approach by reference to citizens' views. This can help foster ongoing support through improving transparency and accountability and facilitate the development of such terms where societal expectations evolve due to innovations or increased

[40] A Faulkner and L Poort, 'Stretching and Challenging the Boundaries of Law: Varieties of Knowledge in Biotechnologies Regulation' (2017) 55 Minerva 209.

[41] Pam Carter, Graeme Laurie, and Mary Dixon-Woods, 'The Social Licence for Research: Why Care.data Ran into Trouble' (2015) 41 Journal of Medical Ethics 404.

familiarity with such innovations. This thesis supports the view that societal knowledge provides insights into the 'context in which both the law or regulatory framework and the new technology will function and develop.'[42]

The thesis' approach emphasises the relevance of the theories outlined in this chapter's introduction, which explore the relationship between law and society in the context of innovative technologies. Some argue that the law struggles to keep pace with innovations[43] due to society's increasingly pluralistic nature, making it more difficult to achieve any consensus, which further undermines law's ability to govern effectively. This thesis argues that perhaps this may be an opportunity rather than a problem. This approach draws from ideas associated with the theory of responsive law,[44] which creates space for ongoing citizen engagement after the passing of legislation. Part I of the thesis noted that certain terms were used and imbued with certain ideals during the legislative process. These included the practice of research, the role of rights in relation to their function in society, societal expectations, and the public interest in research. This thesis supports the position that where norms are used to determine what is lawful, citizens should help develop such laws. This justifies the role and relevance of citizens at the heart of this thesis. These ideas align with Nonet and Selznick's work on responsive law, [45] which argues that legislation developed through

[42] Faulkner and Poort (n 40).

[43] Lyria Bennett Moses, 'Recurring Dilemmas: The Law's Race to Keep up with Technological Change' [2007] SSRN Electronic Journal <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=979861#>.

[44] Nonet and Selznick (n 2).

[45] Nonet and Selznick (n 2).

dialogue with society responds better to societal needs. The approach thus focuses upon establishing a common good or upholding common values and encourages the revision of rules to reflect developments in these values better. The focus during the legislative focus should thus focus on the problem itself.

To fully understand the importance and distinctiveness of responsive law, Nonet and Selznick's work proposes a typology between three ideal types of law: repressive, autonomous, and responsive. Repressive law operates as an instrument for the imposition of power through a top-down model where the point of law is not to secure rights but to attain the ruler's objectives.[46] Such legal orders struggle to preserve its aura of legality where accountability is demanded.[47] They can co-exist within a legal order, but their logics are different with the variables of 'the role of coercion in law, the interplay of law and politics; the relation of law to the state and to moral order…civic participation; legitimacy; and the conditions of obedience'[48] vary. Autonomous law is identified with the rule of law, where stress is placed on the centrality of rules as the basis for decision-making, based on regularity and fairness rather than substantive justice.[49]

Nevertheless, there are tensions between formal and substantive forms of justice where the 'justice of autonomous law is experienced as sham and arbitrary when it frustrates the very expectations of fairness it has encouraged.'[50] Nonet

[46] Martin Krygier, *Philip Selznick: Ideals in the World* (Stanford Law Books 2012).

[47] Nonet and Selznick (n 2), 57.

[48] Ibid, 14.

[49] Nonet and Selznick (n 2).

[50] Ibid, 67.

and Selznick claim that formal legal justice alone is insufficient as it 'tends to serve the status quo.'[51] Furthermore, a

> *rigid adherence to precedent and the mechanical application of rules hampers the capacity of the legal system either to take new interests and circumstances into account or to remedy the effects of social inequality.*[52]

Responsive law arises in light of increased social complexity and rapid changes, which push towards generalising the law's objectives.[53] The notion of 'autonomous law' may be more beneficial over 'repressive law.' Nevertheless, Nonet and Selznick argue that the formalistic justice that can be secured by autonomous law is not the highest form of justice that a legal system should seek to secure. As a result, they argue that 'the tension between procedural and substantive justice generates forces that push the legal order beyond the limits of autonomous law.'[54]

One of the defining characteristics of responsive law is the recognised existence of implicit values within rules and policies.[55] Rules no longer operate as exhaustive reasons for decisions but are 'treated as guides to purposes' and 'grounds of criticism, as rules and decisions come to be evaluated according to underlying purposes.'[56] This shift is highly relevant to this thesis as the 'erosion

---

[51] Ibid, 53.

[52] Ibid, 53.

[53] Ibid, 79.

[54] Ibid, 67.

[55] Ibid, 80.

[56] Krygier (n 46), 179.

of the strength of authority' creates a blurring between legal logic and the logic of the wider world. This blurring 'opens the legal process to more voices because the lawyer's voice has lost its unique authoritativeness.'[57] Therefore the law becomes a problem-solving, facilitative enterprise. It can bring to bear a 'variety of powers and mobilise an array of intellectual and organisational resources.'[58] Whilst responsive law encourages civic participation, this is a risky endeavour as lawmaking institutions become more vulnerable to power imbalances in society and more readily focussed on a narrow range of special concerns.[59] Therefore, careful attention must be paid to how the 'virtues of openness, flexibility and adaptability' do not degenerate into opportunism.[60] Responsive law presumes that the purposes that guide and shape laws can be made sufficiently objective and authoritative to control adaptive rule-making and guard against such opportunism.

Fuller's argument[61] that legislation's effectiveness depends upon its respect for the principles of 'reciprocity' and 'co-operative effort' is of relevance here. The principle of reciprocity aligns to the traditional 'vertical' interactions that keep a society intact as, in a large society, 'members' interests are too disparate to sustain communal enterprises.'[62] This links to Fuller's 'morality of duty,' which lays down

[57] Ibid, 180

[58] Nonet and Selznick (n 2).

[59] Ibid.

[60] Krygier (n 46), 192.

[61] Fuller and Winston (n 3).

[62] Ibid, 67.

the basic rules 'without which an ordered society directed toward certain specific goals must fail of its mark.'[63] However, the principle of co-operative effort speaks to horizontal relationships between members, based on the principle of shared commitment, understood in terms of common ends or, perhaps, the public interest. This shared commitment relates to Fuller's notion of the 'morality of aspiration' that incorporates cherished values. A combination of both forms of interaction must be present in the law.

Fuller's views align with Van der Burg and Frans Brom's[64] arguments that introducing a statutory norm is no longer sufficient in modern societies to secure compliance. Instead, compliance relies on some form of cooperation of citizens. Without this, the effectiveness of such laws is 'doubtful.'[65] The context of modern data processing practices, however, challenges the idea that there are indisputable values that can provide guidance. Van der Burg and Frans Brom noted, '[i]n response to these social and technological changes, our positive morality and ethical theories have to develop as well.' [66] The importance is suggested by the perceived change in the character of law, towards more open standards and procedural norms and the use of ethics codes.[67]

---

[63] Lon Fuller, *The Morality of Law,* (Yale University Press 1969), 5.

[64] Wibren Van der Burg and Frans Brom, 'Legislation on Ethical Issues: Towards an Interactive Paradigm' (2000) 3 Ethical Theory and Moral Practice 57.

[65] Ibid, 60.

[66] Ibid.

[67] Ibid.

These theories have been combined to create a theory of an 'interactive legislative' approach. This approach emphasises the need for a process of interaction between the legislature and society, building on Fuller's notion of 'co-operative effort.'[68] In emphasising the horizontal level of interaction between members, such a practice should foster further societal acceptance by incorporating their ideas, enabling the legislature to understand existing practices better and facilitate expert input. The interactive legislative approach emphasises the need for interaction using open norms that can continuously develop during implementation. As part of this, it does not presuppose consensus but utilises it as an aim, justifying the continuous process of norm development. Whilst consensus may not always be possible, even a crystallisation of the differences that characterise the issues at stake helps the process of norm development within the law. This demonstrates that there is still a role for societal expectations in developing the law in this area.

Palazzo and Scherer argue that where conflicts cannot be resolved by reference to a shared background of values, as may be the case in the governance of innovative and unfamiliar technologies, communication operates as the source of interaction and mutual recognition.[69] Poort, drawing upon the arguments of Vorstenbosch and Ippel, identified four ways this input improves the quality of laws.[70] Firstly, civic engagement in the development of norms can foster their

[68] Fuller (n 63).

[69] Guido Palazzo and Andreas Georg Scherer, 'Corporate Legitimacy as Deliberation: A Communicative Framework' (2006) 66 Journal of Business Ethics 71.

[70] Poort (n 1).

acceptance. Secondly, while citizens may have a low awareness of some current innovations, this may not always be the case. Expectations and views may be 'concretised' in response to certain events, suggesting a need to 'leave room in the legal framework to substantiate the legal rules in concrete situations.' [71] Furthermore, the interactions engendered by public engagement can enable the legislature to develop a more comprehensive understanding of the current landscape of norms and expectations. Finally, it recognises the limits of actors' expertise in creating legislation, enabling the participation of other forms of knowledge and expertise. The approach encourages communication, interaction, and coordination to develop a more effective and legitimate law. Its applicability is highly relevant in the context of this thesis, where the principles and values held by individuals within society remain unclear, and the facilitation of ongoing reflection and debate appears worthwhile.[72]

A focus on communication and interaction aligns with more deliberative conceptions of democracy. Here, public deliberation is at the heart of political decision-making, and the focus of legitimacy is on these 'inputs.' The legitimacy of such decision-making derives from their ability to reflect a plurality of competing discourses and is predicated on the idea that this process of deliberation reinforces a sense of society.[73] Doing so reconceives the law as a facilitator of responses to social needs, anticipating and adapting to existing and future needs. Such

[71] Ibid, 11.

[72] Ibid.

[73] Ibid.

developments would occur in a manner that recognises these needs rather than by reference to established patterns of thought.[74] As Tam notes, 'any government that does not seriously engage citizens in developing a shared approach will rightly lose public trust and legitimacy.'[75] While this makes it increasingly challenging to distinguish law from other forms of decision-making, it also helps prevent officials from retreating behind specific rules,[76] such as those governing the reuse of administrative data in research. These processes for including individuals must not just concern themselves with the procedure but also the substantive outcomes. As Nonet and Selznick argue, the law should be committed to achieving substantive justice. Consequently, legal institutions become instruments of social change premised upon general goals and implicit values that shape rules and policies.[77]

This recognition of the complexities involved and a shift away from rigid adherence to rules, responsive law creates a sense of civility informed by an 'ethic of responsibility' whereby one considers multiple interests and values.[78] This encourages civility by clarifying the values at stake and being more receptive to diversity.[79] This entails a civil regard for multiple ends and thus encourages

[74] TT Arvind and Aisling M McMahon, 'Responsiveness and the Role of Rights in Medical Law: Lessons from Montgomery' (2020) 28 Medical Law Review 445.

[75] Tam (n 38).

[76] Nonet and Selznick (n 2), 83.

[77] Nonet and Selznick (n 2), 74.

[78] Nonet and Selznick (n 2), 91.

[79] Nonet and Selznick (n 2), 91.

negotiation rather than subordination.[80] This resonates with the literature on the concept of a social licence, which can sustain and legitimise a practice where there are divergences between what is legally permissible and societally acceptable.[81]

With the importance of societal expectations emphasised and a role for their further inclusion found, what remains is to investigate societal expectations further. The analysis in Part I indicated that public engagement before legislative action might be insufficient to ensure societal support's creation and continued existence. I agree with Faulkner's view that legislation has a 'condensed and directive quality that leads to a measure of lock-in.'[82] It has force, but how much? If it has insufficient force, what then? Do we hope nothing high profile happens, such as another Edward Snowden-style expose or Cambridge Analytica-type scandal to shine a spotlight on these practices?[83] As Suchman notes, as legitimacy is a perception, an organisation may diverge 'dramatically from societal norms yet retain legitimacy because the divergences go unnoticed.'[84] Do we hope societal attitudes are sufficiently close to the views and attitudes of the more powerful actors in the legislative process by the time this comes around? Perhaps this is already the case? These questions can be examined by exploring societal attitudes and expectations, which forms the focus of the next section.

---

[80] Nonet and Selznick (n 2), 93.

[81] Aitken and others (n 14).

[82] Alex Faulkner, 'Law's Performativities: Shaping the Emergence of Regenerative Medicine through European Union Legislation' (2012) 42 Social Studies of Science 753.

[83] Chris Harris and Lauren Chadwick, 'Edward Snowden Talks Data Exploitation at Web Summit' *euronews* (November 5, 2019) <https://www.euronews.com/2019/11/03/web-summit-2019-whistleblowers-edward-snowden-and-brittany-kaiser-set-to-speak-at-lisbon-c>.

[84] Suchman (n 29), 574.

5.3 BUILDING BRIDGES BETWEEN THE LAW AND SOCIETAL EXPECTATIONS: THE EMPIRICAL APPROACH

Part II of the thesis concerns the empirical research, with the rest of this chapter outlining and justifying the empirical approach. Annex A provides further detail on the methods used in the thesis. Annex B provides a copy of the online survey completed by respondents. Annex C provides a copy of the questioning route used in the online focus groups.

A mixed methods study was chosen to answer the differences between legal approaches and societal expectations towards reusing administrative data in social science research. The primary aim of the empirical work is to gain a better insight into participants' views and expectations towards this reuse of data. The concept of mixed methods research defies a universal definition due to variations in its use within the literature. Creswell and Piano Clark[85] note there are variances as to what is mixed, including methods and methodologies, where the mixing occurred in the research process, the scope of mixing, and its purpose. Johnson and others' definition of mixed methods is used, as something occurring when a researcher

> *combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration.*[86]

[85] John Creswell and Vicki Plano Clark, *Designing and Conducting Mixed Methods Research* (3rd edn., Sage 2018).

[86] R Burke Johnson, Anthony J Onwuegbuzie and Lisa A Turner, 'Toward a Definition of Mixed Methods Research' (2007) 1 Journal of Mixed Methods Research 112, 123.

A mixed methods study, encompassing both quantitative and qualitative approaches, was considered the most appropriate means for responding to the research questions within this thesis. Quantitative approaches can help highlight areas of different viewpoints. These areas can merit further exploration through a method that can draw out the underlying reasonings used by people when forming their views. Qualitative approaches are critical to exploring a range of insights, enabling a nuanced view of the social realities as understood by participants. It allows the rich exploration of the beliefs and views that can emerge as part of a qualitative study and enables participants to provide new information. In this thesis, the qualitative element follows the quantitative element. The qualitative element is prioritised due to its better alignment with the methods used in Part I and its ability to draw out richer understandings of participants' views.

There have been many reasons proffered for the use of mixed methods research. One such reason is the possibility of using multiple methods to study the same phenomenon to 'cancel out' the bias inherent in a particular method. Using mixed methods triangulation can generate findings that converge, be inconsistent, and contradictory. It is not necessarily the case that convergences are the only valid output as all outcomes mean 'the researcher can construct superior explanations of the observed social phenomena.'[87] As noted by Hesse-Biber,[88] contradictory results can still generate new knowledge rather than signalling a

[87] Ibid, 115.

[88] Sharlene Hesse-Biber, 'Qualitative Approaches to Mixed Methods Practice' (2010) 16 Qualitative Inquiry 455.

failure on the researcher's part. Mixed methods research does not just promote triangulation as combining data and the data findings can complement one another. This combination allows one method to elaborate or clarify results. They can also develop studies by informing the approach of another method. This approach aligns with Creswell and Piano Clark's reasoning that mixed methods research may be justified where one data source alone is insufficient to respond to the research questions. [89]

A wealth of the reasons that justify mixed methods research apply to this thesis. Firstly, quantitative and qualitative research methods have their own weaknesses, so combining them can help offset their respective weaknesses. Here, surveys allow for a larger number of participants' views to be gained, providing certain robustness to these initial findings. Nevertheless, they do not permit the depth of understanding fostered through qualitative methods, such as focus groups, where participants can reflect, explore, and further discuss their views. A mixed methods study also facilitates 'completeness,' understood as a more comprehensive account than would be provided through a single method.[90] Here, the quantitative aspect can help highlight potential areas of divergence between the legal approach and societal expectations, but the qualitative element provides space for further elaboration and exploration of the context in which participants determine their responses. Where the quantitative research phase identifies

[89] Creswell and Plano Clark (n 85).

[90] Alan Bryman, *Mixed Methods* (Sage 2006).

divergences, the qualitative phase can provide additional insights and a richer understanding of some of the trends identified in the survey.

The quantitative strand of this empirical work consisted of a survey conducted between October and December 2020 and completed by 302 participants. The survey was open to all UK residents over the age of 18 and, due to the restrictions in place, took place online. Chapter 6 presents the quantitative findings. The qualitative strand, consisting of 5 online focus groups, took place in May 2021. Thirty-six participants agreed to take part and were conducted online through Zoom due to the restrictions in place. Using online methods may have excluded certain people from participating in the study. However, the circumstances left no other viable option to bring together people to debate and discuss the issues explored within this thesis. Chapters 7 and 8 analyse the qualitative findings.

The findings from the quantitative research helped inform the approach of the qualitative aspect in two ways. It highlighted potential areas where respondents' views towards the appropriateness of certain aspects of the reuse of administrative data in research diverged from the GDPR and DEA's legislative approaches. Furthermore, open questions within the online survey enabled respondents to provide their views in their own words, offering fresh insights. The existing literature and quantitative findings helped to inform the questioning route used in the focus groups. The mixed methods approach highlights areas where the qualitative and quantitative data do not align, enabling a reflection upon why this may have occurred. As noted previously, a lack of convergence between the two strands is not necessarily a methodological failure and can

provide valuable insights into how participants form and shape their views. The next section of this chapter presents an overview of the empirical approach towards investigating societal expectations regarding reuses of administrative data in social science research.

5.3.1 Exploring Societal Expectations Through a Survey

Surveys are an established method to collect inputs from a sample of the public. Whilst they can be a standalone method, the results inform the next stage of the empirical research in this thesis. Surveys can permit views to be gathered at scale, thus facilitating an investigation of the responses of certain sub-groups.

Some may perceive an incongruence between the legal analysis and a survey, due to the former's foundations in constructivist tenets. To some, this combination creates inherent friction within the research. Surveys are traditionally understood to be underpinned by a positivist or post-positivist position. Nevertheless, some elements of a post-positivist worldview informed the survey design. Most notably, the theory that divergences existed between legal approaches and societal expectations towards the reuse of data was tested through variables drawn from existing research. By removing the ability of respondents to contribute their constructions of the meaning of terms within survey questions, surveys appear to undermine this by preferring a more 'neutral' rendition of reality.[91]

[91] Norma RA Romm, 'Employing Questionnaires in Terms of a Constructivist Epistemological Stance: Reconsidering Researchers' Involvement in the Unfolding of Social Life' (2013) 12 International Journal of Qualitative Methods 652.

Nevertheless, it is important to stress two points. Firstly, the use of open questions within the survey already permits respondents to contribute their own constructions, meanings, and understandings to certain questions. The apparent tension between the methodological choices may be eased as the survey also raised respondents' attention to topics not customarily considered easy to raise. Whilst others have used this in 'taboo' contexts, I believe this reasoning could be applied here.[92] The 'openings' created by this survey, through introducing topics and demonstrating to participants that they did hold a view, serve to spark the possibility of reconsidering the issues raised. This chimes with the views of Room, who suggests a constructivist epistemological stance allows 'one to consider both closed- and open-ended questions as opportunities for enabling people to develop their opinions as they engage with the questionnaire.'[93] Secondly, the intention was always to complement the findings from the more quantitatively oriented work with a further, qualitatively oriented approach. The latter operates to explore the socially constructed nature of expectations within the context of approaches towards reusing administrative data in research. By acknowledging that the survey results are open to subsequent reconstruction, I seek to bring this methodology closer to a constructivist-oriented approach.

The explanations for respondents' views presented in Chapter 6 remain somewhat cautious and provisional. There may be multiple interpretations of what participants might have expressed in their responses. This aligns with

[92] Ibid.

[93] Ibid.

acknowledgements that there is no 'royal road to ultimate knowledge,'[94] creating space for both quantitative and qualitative methods to explore phenomena. The thesis chimes with approaches that support qualitative methods as the dominant method in a study yet recognise the benefits of additional quantitative data. The quantitative aspect of the empirical work is to go further than provide statistical findings, though these remain informative. Instead, it highlights some of the factors that appear to be relevant to societal expectations, enabling the qualitative research to delve beneath the surface of these findings. The quantitative data provides a starting point for exploring the research questions of this thesis, but the purpose of the empirical work is to develop an in-depth understanding. Qualitative methods better facilitate this aim.

Thirdly, the survey allowed participants to consent to be recontacted with information about further empirical work. As this second phase of the empirical work is based upon the findings from the survey, it allows those participants an opportunity to shed further light on the views they provided. Further detail of the second phase of the empirical work forms the substance of the next section.

5.3.2 Exploring Societal Expectations Through Focus Groups

As well as surveys, this thesis explores societal views and expectations through focus groups to enable a more in-depth consideration of the issues explored in this thesis. The aim is to complement and build upon the findings in the survey to capture nuances and the diversity of views held by participants. This is particularly useful in exploring complex and controversial topics, which may

[94] Manfred Bergman, *Advances in Mixed Methods Research* (SAGE Publications Ltd 2008), 4.

describe certain aspects of the data processing explored within this thesis. The focus groups explore expectations and attitudes towards reusing administrative data in social science research, particularly through facilitating discussion and debate between participants. This can provide insights on their constructions of the key terms found within data protection law. The focus groups also help explain some of the quantitative findings through a deeper exploration of the findings.

Qualitative methods effectively enable the collection of rich data on different viewpoints and perspectives that emerge from discussions and debates between participants.[95] Focus groups generate interactions between participants, making them useful for exploring opinions, ideas, and views. [96] Through discussions and debates, focus groups can generate insights into what people think, how they reach their views, and why they think a certain way.[97] Focus groups generate rich data on participants' thoughts through complementary and argumentative interactions in a way that is only accessible through group interactions.[98] Furthermore, focusing on the debate between participants enables the researcher to minimise their input, facilitating participants to define their key concepts and use their experiences. It gives participants greater control over the process and can thus reduce the power imbalance between the researched and the researcher, minimising the latter's 'authoritative voice.'[99]

---

[95] EA Whitley, N Kanellopoulou and J Kaye, 'Consent and Research Governance in Biobanks: Evidence from Focus Groups with Medical Researchers' (2012) 15 Public Health Genomics 232.

[96] Ibid.

[97] Jenny Kitzinger, 'Qualitative Research: Introducing Focus Groups' (1995) 311 BMJ 299.

[98] David Morgan, *Focus Groups as Qualitative Research* (Sage Publications 2009).

[99] Pranee Liamputtong, *Focus Group Methodology: Principles and Practice* (Sage 2012).

The use of scenarios and open-ended questions helps the researcher to understand what is important to participants, in their own words, through their own experiences. It responds to the problem that people's 'people's knowledge and attitudes are not entirely encapsulated in reasoned responses to direct questions.'[100] It allows the researcher to draw insights from the plethora of communication forms people use in their daily lives; anecdotes, jokes, and even non-verbal kinds such as facial expressions and body language.[101] These forms of communication enable other participants to respond and build on these reactions to reveal areas of agreement, conflict and uncertainty.[102] Moreover, focus groups allow researchers to observe how opinions and views may alter during a session and the arguments and points that influenced these changes. This method can help highlight societal values or group norms through interpersonal communication.[103]

Nevertheless, focus groups do not seek to find consensus, which is not the aim of this thesis. A consensus is not presupposed, and differences in participants' expectations and views are valuable insights in themselves. A range of responses provides a greater understanding of the attitudes and perceptions of participants.[104] Focus groups align well with the method used in Part I as they help uncover multiple interpretations of the key terms and approaches found

[100] Kitzinger (n 97).

[101] Ibid.

[102] Liamputtong (n 99).

[103] Kitzinger (n 97).

[104] Liamputtong (n 99).

within the legislation. Just like the actors in the legislative processes analysed in Part I, focus groups provide opportunities to interact and explore the socially constructed understandings that participants offer. These multiple understandings and meanings can be further analysed to explore how participants understand and negotiate differences of views. This can generate findings on why participants' views may differ from the legal approach through the collective efforts to debate and discuss data reuse.

The focus group findings provide a richer understanding of the differences between legal approaches and participants' expectations of reusing administrative data in social science research. Due to the small number of participants in the qualitative aspect of the empirical work, the findings cannot be generalised to a wider population. Instead, they provide an insight into some of how certain people approach this reuse of data. These differences can be used as a springboard to explore potential future steps that can help promote the legitimacy of data protection law in such a way as to foster opportunities without infringing other interests and rights. Part III of the thesis will explore these issues after presenting the empirical findings in Part II.

6 INVESTIGATING RESPONDENTS' VIEWS TOWARDS THE ACCEPTABILITY OF SECONDARY USES OF ADMINISTRATIVE DATA IN RESEARCH THROUGH A SURVEY

6.1 INTRODUCTION

The analysis in Part I suggested that the legislative approaches referred to the expectations and desires of citizens and the potential to secure societal goods to justify their respective approaches. The outcomes of research were constructed as providing a robust evidence base for better policymaking and a means of improving citizens' lives. Chapter 1 outlined a range of data-sharing initiatives that have sought to capitalise upon this construction of secondary uses of government-held data as an activity in the public interest.

There is increasing recognition of the importance of societal support for secondary uses of data and the need to better respond to peoples' concerns. Empirical research has explored the importance of citizens' views towards secondary uses of their data for research purposes. However, it often focuses on secondary uses of health data for medical research purposes.[1] These highlight areas that are important to consider in the context of secondary uses of administrative data in social science research, including the role of consent,[2] the

[1] 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 <https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-016-0153-x#Fn72>.

[2] Susan Brown Trinidad and others, 'Informed Consent in Genome-Scale Research: What Do Prospective Participants Think?' (2012) 3 AJOB Primary Research 3, and Nandini Saxena and others, 'Understanding Canadians' Attitudes and Expectations Citizens' Dialogue on Privacy and the Use of Personal Information for Health Research in Canada Public Involvement Network' (2006) <http://cprn3.library.carleton.ca/documents/43383_en.pdf>.

influence of respondents' attitudes towards actors involved in data-sharing,[3] and the impact of respondents' views towards the risks raised by secondary uses of data.[4]

Nevertheless, Nissenbaum posits that the context in which data are shared influences peoples' views on its acceptability.[5] Therefore, it cannot be assumed that the findings from these studies can be extrapolated to other types of research which use different kinds of data. The research findings of Cameron and others,[6] which explicitly explores reuses of administrative data in social science research, support distinguishing this data processing from research exploring medical research. This research found that many attributed a low value to social science research, describing its findings as 'common sense.'[7] Furthermore, research has suggested that people have higher trust levels in the NHS than the British government and academic institutions.[8]

---

[3] Amy L McGuire and others, 'DNA Data Sharing: Research Participants' Perspectives' (2008) 10 Genetics in Medicine 46 and TNS BMR, 'Open Data Dialogue: Final Report' (2013) <https://www.force11.org/presentation/open-data-dialogue-final-report>.

[4] Susan Brown Trinidad and others, 'Genomic Research and Wide Data Sharing: Views of Prospective Participants' (2010) 12 Genetics in Medicine 486.

[5] Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press 2009).

[6] Daniel Cameron, Sarah Pope and Michael Clemence, 'Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes' (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/>.

[7] Ibid, 16.

[8] Ipsos MORI, 'Public Attitudes to the Use and Sharing of Their Data' (2016) <https://www.slideshare.net/IpsosMORI/public-attitudes-to-the-use-and-sharing-of-their-data>, 3.

Much of the existing empirical research has been conducted upon informed individuals. Some empirical work has recruited those with specific health conditions.[9] Furthermore, most people have had direct interactions with medical professionals and have been a 'patient' of one description or another at some point. These experiences give people a frame of reference through their interactions with the NHS to develop their views. Cameron and others'[10] study sought to inform participants about secondary uses of administrative data in research that elevated their level of knowledge far beyond that of most citizens. The workshops undertaken as part of that research lasted six hours and included presentations by researchers. Therefore, the views and attitudes of respondents towards reusing administrative data in social science research who have not been provided with significant information on this data processing practice remain unclear.

The results of a survey that aimed to investigate respondents' views towards reusing administrative data in social science research that would reflect citizens' knowledge levels more closely are reported here. The primary aim was to understand what respondents' feel to be acceptable secondary uses of their administrative data in social science research. Therefore, this quantitative analysis can be used to compare respondents' views to the legislative approaches identified in Part I.

This survey considers respondents' views towards the acceptability of five factors; (i) the reuse of different types of administrative data, (ii) the level of

[9] McGuire and others (n 3).

[10] Cameron and others (n 6).

identifiability of administrative data, (iii) the actors involved in reusing administrative data in research, (iv) the purpose for which administrative data may be reused in research, and (v) the views towards the acceptability of linking different types of administrative data.

The findings in this chapter provide insights on the acceptability of different aspects of reusing administrative data in research. The findings suggest that most respondents felt the reuse of anonymous data to be acceptable in this context. This approach diverges from what is permissible under the General Data Protection Regulation (GDPR)[11] and the Digital Economy Act 2017 (DEA). Furthermore, the findings also indicate that respondents had different views towards the acceptability of different actors being involved in reuses of administrative data in research, which is not reflected in the legislative approaches. The findings highlight the diverse range of views held by the respondents, which potentially challenges the law's ability to demonstrate its legitimacy. Within this thesis, legitimacy is premised upon permitting actions that are assumed to be acceptable based upon a socially constructed system of norms, beliefs, views, and definitions. How can legitimacy be demonstrated if no single system exists? A reflection of previous empirical work enables further insights into the aspects of secondary uses of administrative data in social science research that may influence citizens' views.

[11] EU General Data Protection Regulation (GDPR): 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.

The conclusion considers how these quantitative findings drive the thesis forward by providing evidence for the qualitative research, presented in the following two chapters. The limitations of the survey findings demonstrate the need for further research that can better shed light on the underlying reasoning that respondents use when determining the acceptability of reusing administrative data in research. Qualitative methods can better facilitate an exploration of what respondents think and the underlying drivers and motivations of these views. A qualitative method, such as focus groups, builds upon the survey findings to gain a deeper, richer insight into some of the underlying justifications and lines of reasoning that individuals use in determining the acceptability of reusing administrative data.

6.1.1 Methods

Annex A describes the methodology of this thesis, including the survey design and dissemination, the respondents' demographic characteristics, and the analytical approach. Annex B provides the survey as accessed by the respondents.

Overall, the survey presented 50 Likert-style items on a five-point scale. These explored respondents' views towards the acceptability of reusing administrative data in social science research. These items were grouped by reference to factors identified in the law and the empirical literature as relevant to either the legal governance of secondary data uses or participants' views in existing studies. Questions four and five included optional follow-ups that were open-ended questions. These were included to enable respondents to discuss their views further or provide explanations and justifications that the survey questions could not collect.

The survey was open to those over the age of 18 and living in the UK and 302 responses were received. As further detailed in Annex A, most respondents were aged below 40 and over 70% identified as female. Furthermore, over 80% of respondents identified and being white and over half the respondents were in full or part-time employment.

6.2 RESULTS

Descriptive statistics (frequencies and percentages) were used to summarise all the variables of interest. Table 1 provides the results on the acceptability of the reuse of different types of administrative data in research. The Likert scale was coded so that Very Acceptable=1 and Very Unacceptable=5, so lower scores indicated higher acceptability. Where respondents selected 'not applicable' to an item, this was coded as missing data and excluded from the analysis to prevent these responses skewing the data. As a Likert scale was used, the variables were treated as ordinal variables and thus non-parametric tests were performed on the data.

6.2.1 The Acceptability of Reusing Different Administrative Data Types in Research.

Table 1 displays the descriptive statistics concerning the acceptability of reusing different data types in research. These types were intended to provide respondents with a better understanding of the range of information that could be categorised as administrative data. Respondents' responses were excluded where they selected 'Not Applicable.' Respondents found it most unacceptable for their address to be reused, expressing this was either very or quite unacceptable (51.3%). However, this means nearly a roughly equal number (48.6%) found it either very acceptable,

quite acceptable, or neither acceptable nor unacceptable. Most respondents (70.6%) found it to be either quite or very acceptable for data relating to the use of public libraries to be reshared.

Table 1: Descriptive Statistics Concerning the Acceptability of Reusing Different Administrative Data Types in Research

	Very Acceptable		Quite Acceptable		Neither Acceptable nor Unacceptable		Quite Unacceptable		Very Unacceptable	
	N	%	N	%	N	%	N	%	N	%
Your name	68	22.8%	66	22.1%	21	7.0%	40	13.4%	103	34.6%
Your address	38	12.9%	68	23.1%	37	12.6%	46	15.6%	105	35.7%
Your date of birth	68	22.9%	111	37.4%	40	13.5%	34	11.4%	44	14.8%
Your ethnicity	102	34.1%	124	41.5%	32	10.7%	16	5.4%	25	8.4%
Your income	52	17.7%	100	34.0%	55	18.7%	45	15.3%	42	14.3%
Your religion	73	25.3%	105	36.3%	46	15.9%	34	11.8%	31	10.7%
Your education history	83	27.9%	110	36.9%	45	15.1%	31	10.4%	29	9.7%
Your employment history	62	20.9%	104	35.1%	45	15.2%	39	13.2%	46	15.5%

Your immigration history	63	23.6%	87	32.6%	40	15.0%	36	13.5%	41	15.4%
The taxes you have paid	67	23.0%	89	30.6%	45	15.5%	40	13.7%	50	17.2%
Information about your access to benefits provided by the UK government	67	23.3%	90	31.3%	44	15.3%	37	12.8%	50	17.4%
Your housing situation e.g. renting or homeowner	75	25.3%	117	39.5%	45	15.2%	21	7.1%	38	12.8%
The vehicles that you own	63	22.0%	87	30.4%	49	17.1%	43	15.0%	44	15.4%
Your use of public transport	102	34.8%	91	31.1%	39	13.3%	29	9.9%	32	10.9%
Your use of public libraries	105	36.0%	101	34.6%	36	12.3%	26	8.9%	24	8.2%
Your use of public parks	98	33.9%	94	32.5%	38	13.1%	26	9.0%	33	11.4%
Your use of Council owned gyms, leisure centres, and facilities	91	31.0%	104	35.4%	48	16.3%	22	7.5%	29	9.9%

Table 2 provides further details on the survey data regarding the acceptability of reusing different administrative data types in research that indicates a significant range of views towards the acceptability of reusing different types of administrative data in research.

Table 2: Descriptive Statistics Concerning the Acceptability of Reusing Different Administrative Data Types in Research

		95% Confidence Interval for Mean					
	Mean	Lower Bound	Upper Bound	Median	Variance	Std. Deviation	Range
Your name	3.19	2.99	3.40	3.00	2.628	1.621	4
Your address	3.36	3.17	3.55	4.00	2.231	1.494	4
Your date of birth	2.58	2.41	2.75	2.00	1.831	1.353	4
Your ethnicity	2.14	1.99	2.29	2.00	1.493	1.222	4
Your income	2.74	2.58	2.91	2.00	1.753	1.324	4
Your religion	2.46	2.29	2.62	2.00	1.662	1.289	4
Your education history	2.37	2.21	2.53	2.00	1.606	1.267	4
Your employment history	2.63	2.46	2.80	2.00	1.813	1.346	4
Your immigration history	2.65	2.48	2.83	2.00	1.930	1.389	4
The taxes you have paid	2.70	2.52	2.88	2.00	2.031	1.425	4
Information about your access to benefits provided by the UK government	2.68	2.50	2.86	2.00	2.044	1.430	4

Your housing situation e.g. renting or homeowner	2.40	2.24	2.56	2.00	1.630	1.277	4
The vehicles that you own	2.65	2.48	2.82	2.00	1.807	1.344	4
Your use of public transport	2.32	2.15	2.48	2.00	1.738	1.318	4
Your use of public libraries	2.18	2.02	2.34	2.00	1.562	1.250	4
Your use of public parks	2.31	2.15	2.48	2.00	1.736	1.318	4
Your use of Council owned gyms, leisure centres, and facilities	2.26	2.11	2.42	2.00	1.567	1.252	4

The diversity of responses, demonstrated in range=4 for all items, supports the view that some respondents' views towards the acceptability of reusing administrative data in research diverge somewhat from what is legally permissible in this context. Despite most respondents finding that it was unacceptable to reuse data pertaining to names and addresses in research, the GDPR does not preclude the reuse of such data. The DEA permits only the reuse of de-identified data, which aligns more closely with the views expressed by respondents, yet it may be possible to share postcodes or surnames that fall within the scope of de-identified data.

6.2.2 The Acceptability of Reusing Data that are More or Less Identifiable in Research

Table 3 demonstrates the responses given towards the acceptability of using more or less identifiable data concerning anonymous data, identifiable data, a definition

of data that corresponds to pseudonymous data, and an item that emphasises the research purpose. The most acceptable form of data to be reused in research was anonymous data, with most respondents (91.7%) finding it either very or quite acceptable. The most unacceptable form of data to be reused in research was identifiable data. Most respondents (70.2%) found the reuse of identifiable information either very or quite unacceptable. The most popular response was finding the reuse of identifiable data unacceptable (42.1%.)

Table 3: Descriptive Statistics Concerning the Acceptability of Reusing Data that is More or Less Identifiable in Research

	Very Acceptable		Quite Acceptable		Acceptable nor		Unacceptable		Unacceptable			Interval for Mean				
	N	%	N	%	N	%	N	%	N	%	Mean	Lower Bound	Upper Bound	Median	Std. Deviation	Range
Anonymous data so that I cannot be identified	224	74.2%	53	17.5%	13	4.3%	5	1.7%	7	2.3%	1.40	1.31	1.50	1.00	0.841	4
Identifiable data so researchers can access data that identifies who I am	17	5.6%	41	13.6%	32	10.6%	85	28.1%	127	42.1%	3.87	3.73	4.02	4.00	1.251	4
Data that strongly reduces the possibility of me being identified, but does not make it impossible	40	13.2%	93	30.8%	38	12.6%	82	27.2%	49	16.2%	3.02	2.87	3.17	3.00	1.328	4
Whatever form of data is most useful for research	15	5.0%	51	16.9%	78	25.8%	69	22.8%	89	29.5%	3.55	3.41	3.69	4.00	1.216	4

Neither the GDPR nor the DEA requires data to be anonymised before it can be reused in research. The legal gateway created by the DEA permits the reuse of de-identified data, but this does not align with the definition of genuinely anonymous data as there is a small risk that individuals may be re-identified. Only the GDPR permits the reuse of identifiable data in research, suggesting that

its approach diverges further from respondents' views towards what is acceptable in this context as this was the most unacceptable option.

6.2.3 The Acceptability of Different Actors Reusing Administrative Data Types in Research

Table 4 provides the results for respondents' views towards the acceptability of different actors reusing administrative data for research purposes. This demonstrates that most respondents (72.5%) found it either very or quite acceptable for central government to reuse administrative data in research. A similar number (69.2%) found it acceptable for local government authorities to do the same.

Table 4: Descriptive Statistics Concerning the Acceptability of Different Actors Reusing Administrative Data Types in Research

	Very Acceptable		Quite Acceptable		Neither Acceptable nor Unacceptable		Quite Unacceptable		Very Unacceptable			95% Confidence Interval for Mean			
	N	%	N	%	N	%	N	%	N	%	Mean	Lower Bound	Upper Bound	Median	Std. Deviation
Central government e.g. HM Revenue & Customs and the Department for Work and Pensions	78	25.8%	141	46.7%	27	8.9%	34	11.3%	22	7.3%	2.27	2.14	2.41	2.00	1.383
Local government e.g. County Council and City Councils	71	23.5%	138	45.7%	40	13.2%	31	10.3%	22	7.3%	2.32	2.19	2.45	2.00	1.335
Academics at UK universities	72	23.8%	109	36.1%	45	14.9%	40	13.2%	36	11.9%	2.53	2.38	2.68	2.00	1.712
Academics at overseas universities	39	12.9%	75	24.8%	51	16.9%	60	19.9%	77	25.5%	3.20	3.04	3.36	3.00	1.949
Charities	25	8.3%	91	30.1%	69	22.8%	53	17.5%	64	21.2%	3.13	2.99	3.28	3.00	1.644
Community Organisations, which aim to benefit certain groups of people in a community but are not charities	26	8.6%	84	27.8%	71	23.5%	64	21.2%	57	18.9%	3.14	3.00	3.28	3.00	1.575
Businesses for their own commercial purposes	10	3.3%	18	6.0%	25	8.3%	77	25.5%	172	57.0%	4.27	4.15	4.39	5.00	1.121
Business when working in partnership with the UK government	12	4.0%	49	16.2%	55	18.2%	87	28.8%	99	32.8%	3.70	3.57	3.84	4.00	1.432

There was a significant difference between respondents' views towards the acceptability of academics reusing administrative data, depending on whether they were based at a UK or overseas institution. Most respondents (59.9%) found

it either very or quite acceptable for academics based at UK universities to reuse administrative data in research. However, respondents' views were quite split over the acceptability of academics at overseas universities reusing administrative data in research. The percentage of respondents that found this acceptable (37.7%) was similar to the percentage that found this unacceptable (45.4%.) However, this demonstrates that more respondents did find this to be an unacceptable reuse of administrative data. Furthermore, for academics at overseas universities the most common response was to find this very unacceptable, which suggests that this question prompted particularly strong feelings amongst most respondents.

Most respondents found it unacceptable for their data to be reused by businesses for their commercial purposes (82.5%.) This level of unacceptability remained relatively high even when the reuse concerned businesses working with the government (61.6%.) For each item range=4, indicating that a broad range of views were held by respondents.

6.2.4 The Acceptability of Linking Different Administrative Data Types in Research

Table 5 demonstrates the results from the question that sought to explore respondents' views towards how acceptable it would be to link together different examples of administrative data. Most respondents found it very unacceptable for their name and address to be linked (Mdn=4) and quite acceptable for the other examples to be linked (Mdn=2). The exceptions to this were the responses towards linking data on one's income, taxes, benefits access, and vehicle information. The results demonstrate quite a diverse range of views expressed towards the acceptability of reusing different data types as for every item range=4.

Table 5: Descriptive Statistics Concerning the Acceptability of Linking Different Administrative Data Types in Research

	Very acceptable for it to be linked with other data		Quite acceptable for it to be linked with other data		Neither acceptable nor unacceptable towards it being linked with other data		Quite unacceptable for it to be linked with other data		Very unacceptable for it to be linked with other data		Mean	95% Confidence Interval for Mean		Median	Std. Deviation
	N	%	N	%	N	%	N	%	N	%		Lower Bound	Upper Bound		
Your name	29	9.6%	47	15.6%	21	7.0%	62	20.5%	143	47.4%	3.80	3.64	3.96	4	1.998
Your address	27	8.9%	47	15.6%	34	11.3%	81	26.8%	113	37.4%	3.68	3.53	3.83	4	1.819
Your date of birth	52	17.2%	102	33.8%	30	9.9%	46	15.2%	72	23.8%	2.95	2.78	3.11	2	2.137
Your ethnicity	75	24.8%	125	41.4%	37	12.3%	33	10.9%	32	10.6%	2.41	2.27	2.55	2	1.598
Your income	55	18.2%	96	31.8%	42	13.9%	55	18.2%	54	17.9%	2.86	2.7	3.01	2.5	1.93
Your religion	63	20.9%	109	36.1%	46	15.2%	43	14.2%	41	13.6%	2.64	2.49	2.79	2	1.754
Your education history	68	22.5%	120	39.7%	40	13.2%	37	12.3%	37	12.3%	2.52	2.37	2.67	2	1.686
Your employment history	53	17.5%	114	37.7%	41	13.6%	53	17.5%	41	13.6%	2.72	2.57	2.87	2	1.725
Your immigration history	52	17.2%	100	33.1%	48	15.9%	45	14.9%	57	18.9%	2.85	2.69	3.01	2	1.908
The taxes you have paid	52	17.2%	90	29.8%	41	13.6%	55	18.2%	64	21.2%	2.96	2.8	3.12	3	2.022
Information about your access to benefits provided by the UK government	47	15.6%	95	31.5%	38	12.6%	55	18.2%	67	22.2%	3.00	2.84	3.16	3	2.013
Your housing situation e.g. renting or homeowner	66	21.9%	106	35.1%	44	14.6%	39	12.9%	47	15.6%	2.65	2.5	2.81	2	1.862
The vehicles that you own	54	17.9%	92	30.5%	40	13.2%	60	19.9%	56	18.5%	2.91	2.75	3.07	3	1.958
Your use of public transport	81	26.8%	109	36.1%	40	13.2%	33	10.9%	39	12.9%	2.47	2.32	2.62	2	1.785
Your use of public libraries	84	27.8%	113	37.4%	39	12.9%	31	10.3%	35	11.6%	2.40	2.26	2.55	2	1.703

6.2.5 The Acceptability of Different Purposes of Reusing Administrative Data in Research

Table 6 provides the descriptive statistics concerning the acceptability of different purposes associated with reusing administrative data in research. The most acceptable purposes were to improve the lives of others (84.4%) and to improve the respondent's life (82.8%.) In contrast, most respondents found it unacceptable for the government to use it for any purpose they wished (76.8.) However, the reuse of data by the government for specified purposes, such as helping see where it could improve public services (85.1%) and save money by making public services more efficient and effective (71.2%), was found to be more acceptable.

Table 6: Descriptive Statistics Concerning the Acceptability of Different Purposes of Reusing Administrative Data in Research

	Very Acceptable		Quite Acceptable		Neither Acceptable nor Unacceptable		Quite Unacceptable		Very Unacceptable		Mean	95% Confidence Interval for Mean		Median	Std. Deviation
	N	%	N	%	N	%	N	%	N	%		Lower Bound	Upper Bound		
To improve my life	122	40.40%	128	42.40%	30	9.90%	9	3.00%	13	4.30%	1.88	1.77	2	2	1
To improve the lives of others	126	41.70%	129	42.70%	23	7.60%	11	3.60%	13	4.30%	1.86	1.75	1.97	2	1.005
To help the government see where it can improve public services	117	38.70%	140	46.40%	25	8.30%	8	2.60%	12	4.00%	1.87	1.76	1.98	2	0.959
To help the government see where it could save money by making public services more effective or efficient	84	27.80%	128	42.40%	48	15.90%	26	8.60%	16	5.30%	2.21	2.09	2.34	2	1.103
To increase our knowledge of the challenges facing society	111	36.80%	121	40.10%	45	14.90%	13	4.30%	12	4.00%	1.99	1.87	2.1	2	1.025
For any purpose the government would like	9	3.00%	21	7.00%	40	13.20%	85	28.10%	147	48.70%	4.13	4.00	4.25	4	1.074

6.3 THE CONSTRUCTION OF RESEARCH AS AN ACTIVITY IN THE PUBLIC INTEREST

The analysis in Part I suggested that the dominant discourses that shaped the GDPR and DEA constructed research as a beneficial activity in the public interest. The focus on its potential to resolve societal problems and capitalise on new technologies to provide novel insights justified a permissive governance approach. The GDPR allowed for derogations and limitations to individuals' rights where data were processed for research purposes. The DEA introduced a specific legal gateway to enable this reuse of data.

As demonstrated in Figures 1 and 2, the most acceptable purposes for administrative data to be reused were to improve respondents' lives, which 82.8% of respondents found to be acceptable, and to improve the lives of others, which 84.4% of respondents found to be acceptable.

Figure 1: Respondents' views towards the acceptability of reusing administrative data in research to improve a respondent's life.

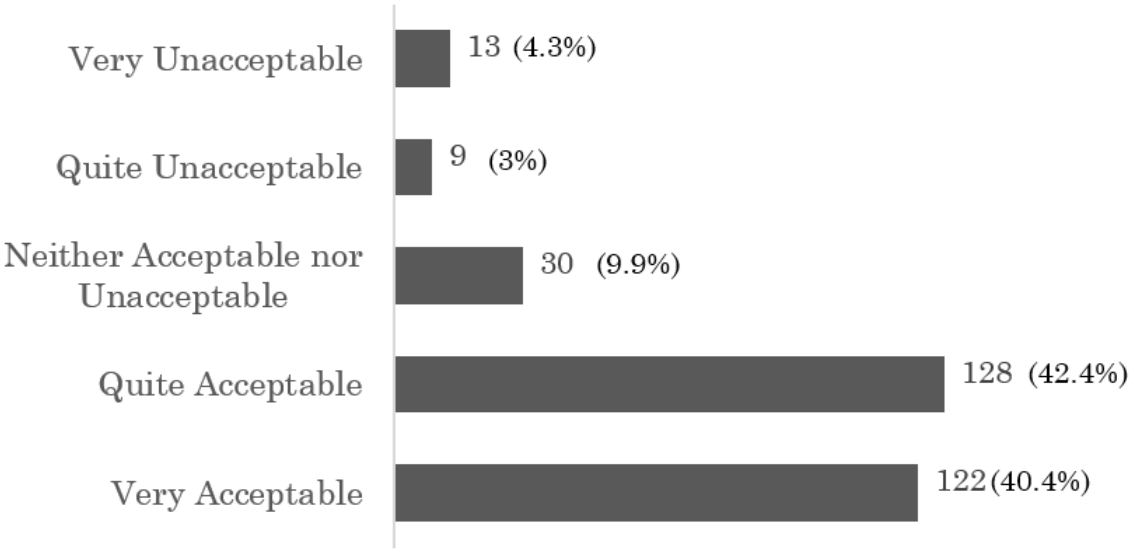
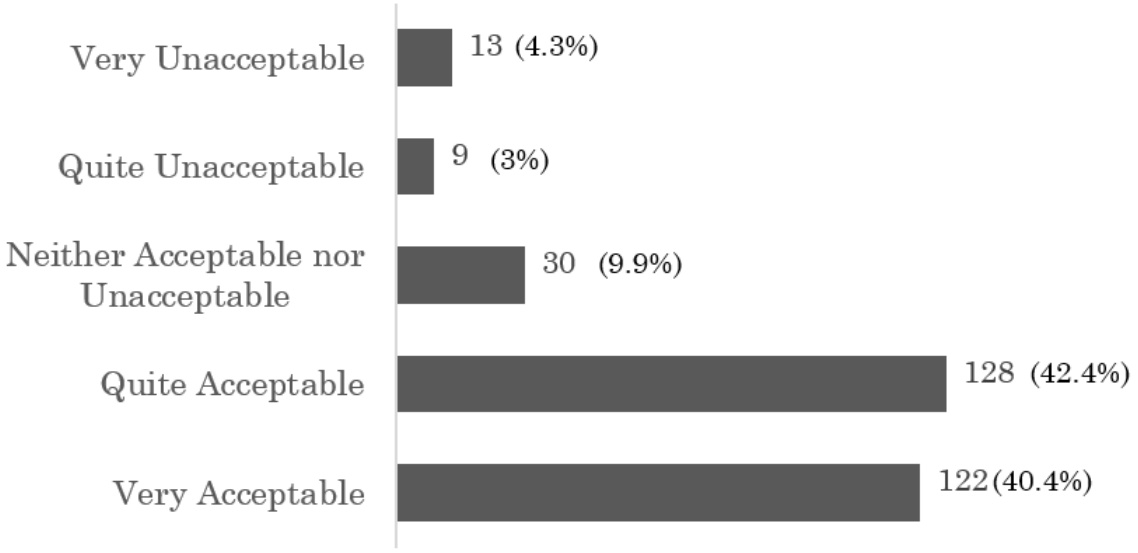
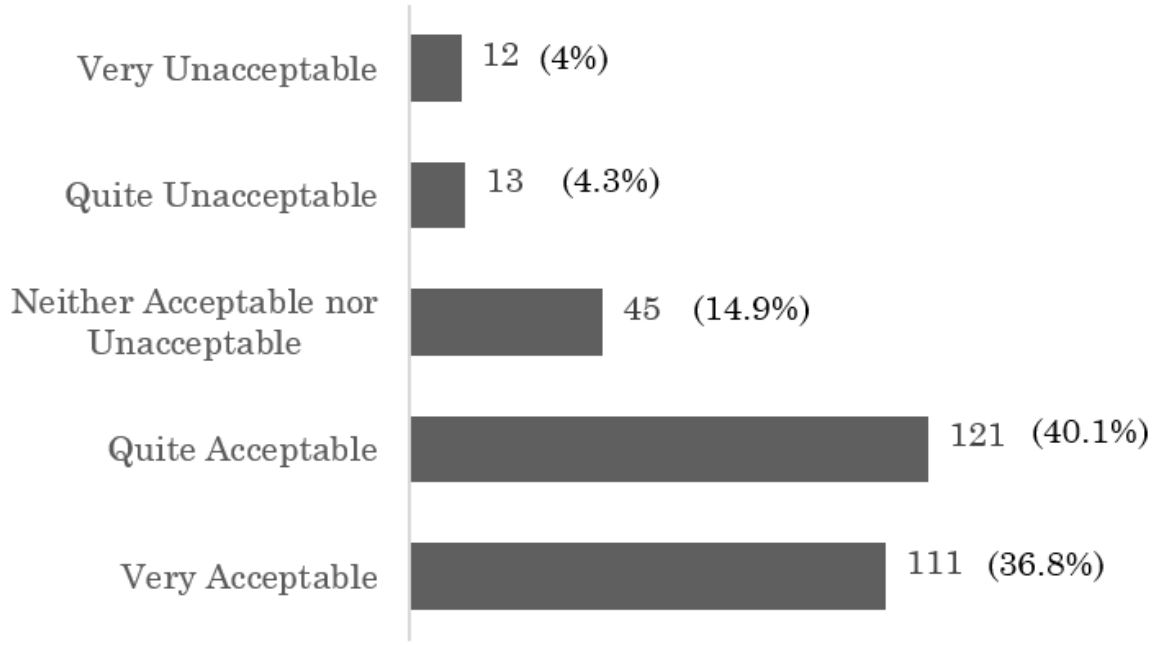


Figure 2: Respondents' views towards the acceptability of reusing administrative data in research to improve the lives of others.



However, Figure 3 indicates that many respondents defined research that warranted access to administrative data more broadly than research providing tangible positive outcomes for individuals in society. Seventy-seven percent of respondents found it either very or quite acceptable for administrative data to be reused to increase our knowledge of the challenges facing society.' This indicates that these respondents found it acceptable for research to be undertaken where the benefit is limited to improving knowledge rather than an explicit improvement in citizens' lives.

Figure 3: Respondents' views towards the acceptability of reusing administrative data in research to increase our knowledge of the challenges facing society.



These findings suggest some alignment between the respondents' views towards the acceptability of reusing data in research where there is no tangible positive outcome and the approach of the GDPR. Part I noted that the GDPR does not require research to produce a particular outcome to benefit from the exemptions to, and derogations from, its provisions. The DEA makes specific reference to the 'public interest' as a criterion that must be fulfilled before researchers may be granted access to administrative data. However, it is unclear whether the 'public interest' in this context would be defined in terms of tangible beneficial outcomes.

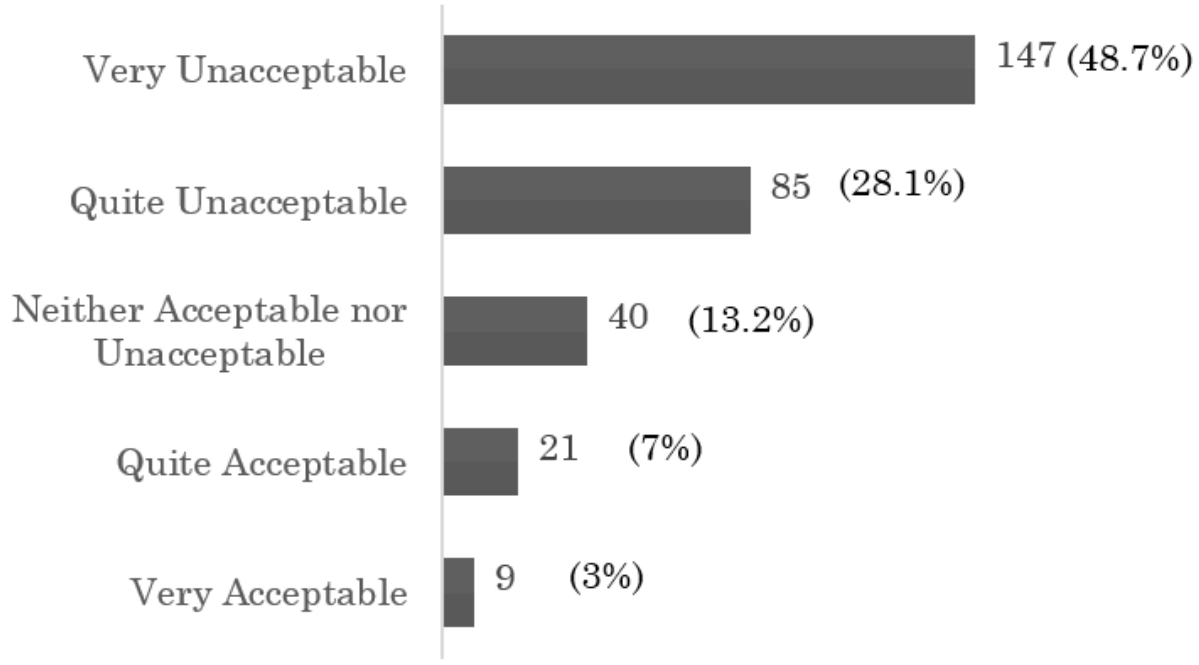
The analysis in Chapter 4 suggested the legislative practices shaping the approach of the DEA attempted to promote the reuse of administrative data in quantitative research that could provide tangible benefits. Doing so suggests the 'public interest' criterion may limit the access of data to research that can

demonstrate its ability to provide tangible outcomes. This suggests a definition of the 'public interest' that may align with research that can provide tangible outcomes. The extent to which it requires such outcomes could be a source of divergence between its approach towards governing acceptable reuses of administrative data in research and respondents' views. This could suggest an area where the DEA is more restrictive in determining acceptable reuses of administrative data in research. Nevertheless, neither the DEA nor the Accreditation Criteria that accompanies the legislative provision explicitly define the public interest by reference to tangible outcomes. Therefore, the DEA does not explicitly preclude research that cannot demonstrate a positive, tangible outcome.

The survey findings suggested that 85.1% of respondents broadly found reuses of administrative data by researchers to inform policy as acceptable. These respondents found the reuse of data in research 'to help the government see where it can improve public services' to be either very or quite acceptable. Most respondents (70.2%) also found it either very or quite acceptable, for data to be used to 'help the government see where it could save money by making public services more effective or efficient.' Nevertheless, this represents an almost 20% decrease in the number of individuals finding it acceptable compared to where data are reused to improve public services.

Figure 4 shows that reusing data for any research purpose the government desired was deemed the most unacceptable by respondents. Here, only 10% of respondents found this purpose acceptable, 28.1% found it quite unacceptable, and 48.7% found it to be very unacceptable. This means that over three-quarters of the respondents (76.8%) found this data reuse to be unacceptable.

Figure 4: Respondents' views towards the reuse of administrative data in research for any purpose that the government would like.



The GDPR's approach suggests that governments would be able to reuse data for any research purpose they liked and still benefit from the exemptions and derogations offered by the GDPR. There are further criteria associated with allowing data to be shared through the gateway created by the DEA. Nevertheless, neither the legislative instrument nor the accompanying documents preclude reuses of data in research desired by the government that are unpopular. The conditions in the GDPR and DEA do not reference popular support for determining the lawfulness of a data processing practice. This suggests there may be a divergence in what the law finds to be acceptable reuses of data and most respondents' views. The quantitative findings suggest that respondents had a more limited definition of acceptable data sharing than any research purpose desired by the government.

This suggests that respondents' views were influenced by the purpose for which administrative data may be used. This approach diverges from the GDPR, which does not require research to be for a particular purpose to benefit from a permissive governance approach. There may also be divergences between respondents' views and the approach of the DEA, for it is unclear how close the criterion of research in the 'public interest' aligns with respondents' constructions of research in the 'public interest.' The findings in Chapter 4 suggested that the dominant discourses that shaped the approach of the DEA considered closer partnerships between academia and industry to be a positive development[12] and suggested a bias towards research that could provide tangible and financial benefits to public authorities.[13] It is unclear how far this construction of research as justifying a permissive governance approach aligns with respondents' understanding of how research reusing administrative data is acceptable. The findings below, which explore respondents' views towards the inclusion of actors with financial and commercial motivations, suggest that respondents may not deem such research to be an acceptable reuse of their administrative data.

6.4 The Construction of Researchers as Acceptable Actors to Reuse Administrative Data in Research

The exceptional approach towards reusing data in research within the GDPR and the DEA's exclusive gateway for accredited researchers suggests these instruments construct researchers' identities as relevant to the governance,

[12] See Chapter 4.2.3.

[13] See Chapter 4.4.2.

including disclosing and reusing data. Here, the legislative instruments suggest a construction of a 'researcher,' which was promoted during their respective legislative processes, as justifying researchers' abilities to reuse data in research. This idea of the 'construction' of a researcher highlights the Foucauldian ideas that underpin Bacchi's framework[14] whereby dominant discourses make certain subject positions available.

The dominant discourses analysed in Part I focussed on the positive attributes of research and the ethical standards upheld by researchers. Doing so rendered negative approaches towards research and researchers nearly unthinkable. Part I demonstrated that there was little explicit distinction made between the different actors that may process data. This section presents the quantitative findings that explore how acceptable respondents found the ability of different actors, such as university researchers, businesses, and the UK government, to reuse administrative data in research. The findings demonstrate a diversity of views, emphasising the need not to treat the 'public' as a homogenous whole. They further suggest a more restrictive construction of 'researchers' as an acceptable actor to reuse administrative data in research.

The respondents were broadly supportive of researchers reusing administrative data in research, but this support was not universal. Figures 5 and 6 demonstrate this breadth, particularly regarding responses to the acceptability of academics at overseas institutions reusing administrative data.

[14] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

Figure 5: Respondents' views towards the acceptability of academics at overseas universities reusing administrative data in research.

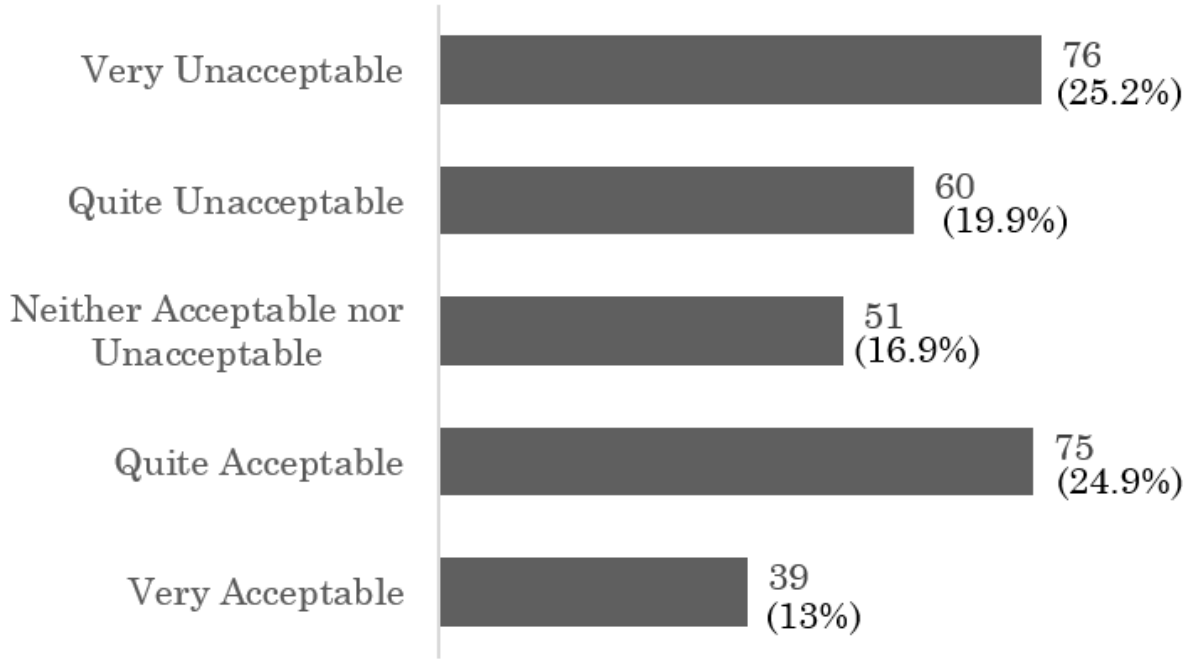
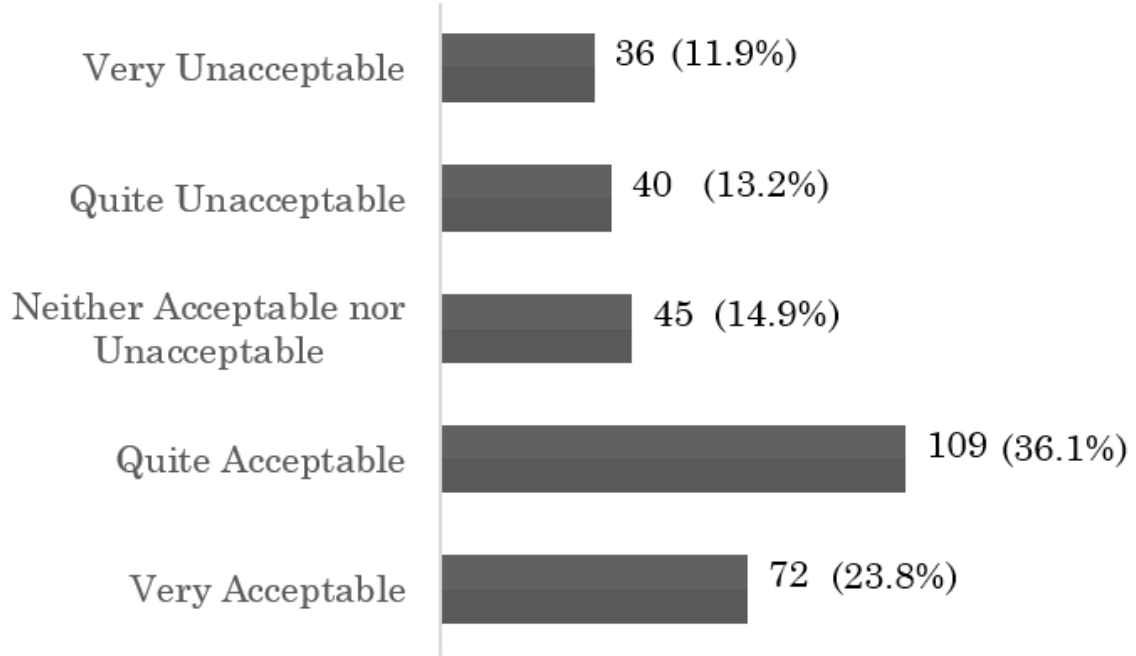


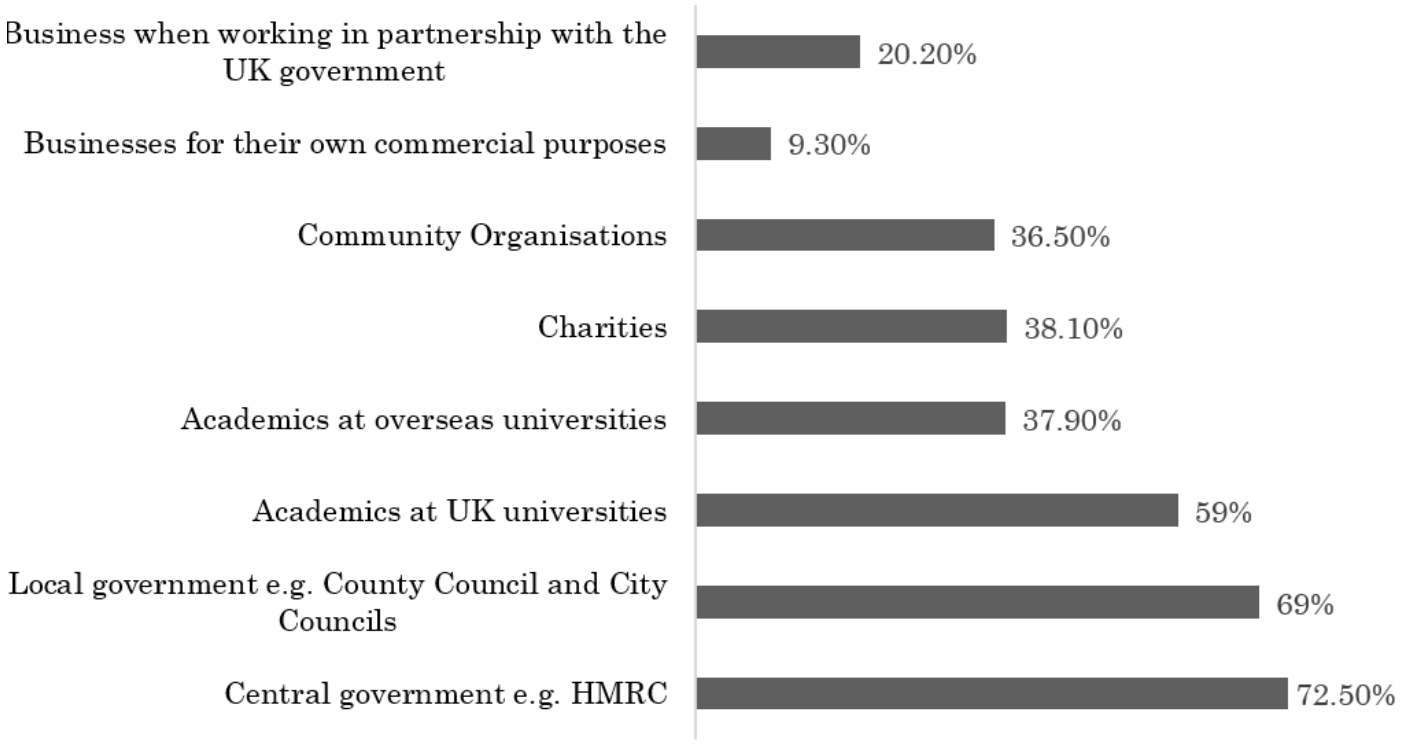
Figure 6: Respondents' views towards the acceptability of academics at UK universities reusing administrative data in research.



Twenty-six percent of respondents found it either very or quite unacceptable for academics at UK universities to have access to administrative data for research purposes. This figure rose to 44.1% of respondents when they were asked about academics at overseas universities. The findings relating to the acceptability of overseas academics reusing administrative data were very diverse. The most common response by respondents was finding this sharing of data very unacceptable (25.2%) followed by it being perceived as quite acceptable (24.9%.) This suggests that respondents were potentially influenced by an interplay of factors that a Likert scale is inapt to explore. For example, perhaps some respondents felt that the acceptability of researchers was intrinsically linked to the purpose of their research. Doing so, they may have felt that those in overseas institutions may not have fulfilled the purposes they deemed acceptable, such as improving the lives of UK citizens or informing UK governmental policies.

A later question asked respondents about the acceptability of reusing data for a range of purposes. However, there was no ability to collect data on how these two aspects of the reuse of administrative data in research interacted to shape respondents' views towards its acceptability. Furthermore, whilst respondents were broadly supportive of university academics reusing administrative data in research, fewer respondents (59.9%) found it acceptable for researchers to reuse administrative data than usage by both central government (70.5%) and local government (69%). Figure 7 provides data on the percentage of responses that found the ability of different actors to reuse administrative data in research as either very or quite acceptable, demonstrating the diversity in views towards the different actors.

Figure 7: Percentage of respondents that found the reuse of administrative data in research by different actors to be either quite or very acceptable.



Commercial entities were the least acceptable actor to be involved in reusing administrative data for research purposes. Neither the GDPR nor DEA explicitly preclude commercial researchers benefitting from their respective permissive governance approaches, yet 82.5% of respondents would find it either quite or very unacceptable for businesses to reuse data for commercial research purposes. The public interest criterion within the DEA's accreditation criteria and need for academic expression within Article 85 GDPR may be interpreted to narrow the scope of researchers that may benefit from the provisions. However, there is no explicit exclusion of these actors. Furthermore, 61.6% of respondents found it either quite or very unacceptable for businesses to reuse administrative

data in research when working in partnership with the UK government. The percentage of individuals who found this data reuse to be either very or quite acceptable more than doubled (9.9% vs 20.2%.) This finding suggests that, through that shift, the purpose and other actors involved influence respondents' perceptions of the acceptability of data reuse. The removal of the emphasis on commercial purpose and the introduction of the relationship with the UK government made this activity, to some, more acceptable. However, it is impossible to tell how far those two changes influenced certain respondents to view the reuses differently from the data. The findings regarding third sector involvement, such as charities and community organisations, were marked by considerable ambivalence. Thirty-eight percent of respondents found it either very or quite acceptable for data to be shared with charities, whilst 39.1% found it either quite or very unacceptable for data to be shared with these organisations.

These findings, alongside the differences in acceptability based upon whether an academic was based at a UK or overseas university, suggest the identity of a data processor as a 'researcher' is insufficient alone for respondents to define the reuse of administrative data in research as acceptable. This suggests that other factors influence respondents' views and thus shape their construction of the 'researcher' as an acceptable actor to reuse administrative data in research. The qualitative responses in the survey support this contention. Several respondents emphasised the interplay of different factors and the broader context in which the reuse of administrative data in research could occur relevant to their assessment of its acceptability. One example of this was the relevance of the purpose for which the data are used. Therefore, to Respondent 68, whilst they

'massively distrust Borders and Immigration with my data' they could 'think of benevolent research even by them.'

The particular unacceptability of businesses reusing data in research, even where explicitly in partnership with the UK government, suggests that commercial aspects of such research significantly influence respondents' views. The qualitative findings collected during the survey that allowed respondents to provide their views towards other organisations not listed in the survey supports this view. Several respondents referenced the possibility of commercial gain as influencing their views. The suggested unacceptability of the involvement of such actors suggests a potential divergence between respondents' views and the legislative approaches. The analysis in Part I suggested that the potential for economic benefits for both public and private entities influenced the legislative approaches of the GDPR and DEA. The approach of the GDPR was heavily influenced by dominant discourses that sought to ensure the protection of personal data did not negatively impact the internal market. Furthermore, the practices that shaped the DEA suggested that a greater interrelationship between industry and academia was a positive development.

6.5 The Acceptability of Reusing Data that can Identify Individuals

The GDPR and DEA take different approaches towards what type of data researchers should be able to reuse. The GDPR permits reuses of personal data, which incorporates identifiable data and pseudonymous data. Furthermore, where personal data are used in research, this research benefits from a presumption of compatibility. This presumption means if one has a ground that justifies the initial

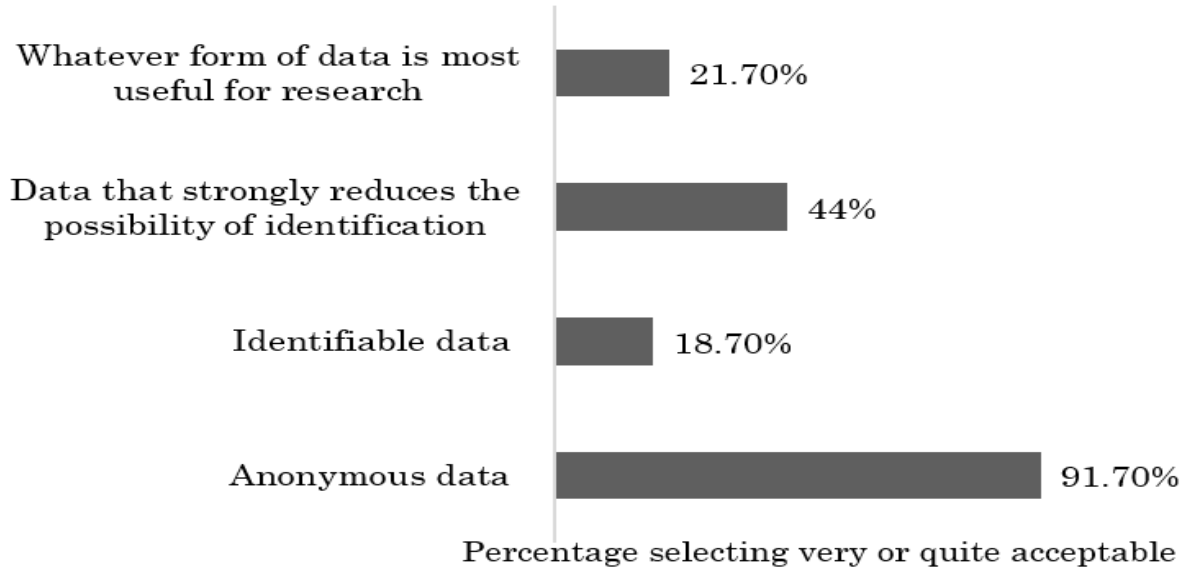
collection of data, its reuse in research is presumed to be justified on that same ground. Truly anonymous data falls outside the scope of the GDPR and therefore can be reused in research without needing to abide by its requirements. The DEA only permits the reuse of de-identified data, which is defined as data that does not disclose a person's identity and it is 'not reasonably likely that the person's identity will be deduced from the information (whether by itself or taken together with other information).' [15] This definition is a less demanding standard than anonymous data, whereby the reidentification of individuals is rendered permanently impossible and suggests some forms of personal data may be reused in research.

The survey sought to explore respondents' attitudes towards these different data types by asking about the acceptability of using anonymous, identifiable, pseudonymised information, and information in the most useful form for research purposes. As 'pseudonymisation' is a very technical term that would be unfamiliar to most respondents, the survey asked about the acceptability of using data that 'strongly reduces the possibility of being identified but does not make it impossible.' This created a spectrum of identifiability that included anonymous, pseudonymous, and identifiable data in a way that was more accessible to respondents.

Figure 8 demonstrates the percentage of respondents that found a particular data type either a very or quite acceptable reuse of administrative data in research.

[15] S64, Digital Economy Act 2017.

Figure 8: Percentage of Responses Finding the Reuse of Different Data Types Either Very or Quite Acceptable.



In contrast to the GDPR's approach, 70.6% of respondents found it unacceptable for researchers to reuse data that could identify individuals for research purposes. The findings indicate that respondents constructed the reuse of pseudonymous information as more acceptable. Nevertheless, the results demonstrate a close split between those that found it acceptable and unacceptable, indicating significant heterogeneity in respondents' views. 44% of respondents found the reuse of such data to be either very or quite acceptable, yet 43.3% found it to be either quite or very unacceptable. Even where the purpose of research was emphasised, by asking for respondents' views towards the acceptability of the reuse data in 'whatever form was most useful for research purposes,' 52.4% of respondents found this unacceptable. The most acceptable form of data to be reused, by a significant extent, was anonymised data. The reuse of anonymous data in research was found to be acceptable by 91.7% of respondents, with 74.1%

of respondents overall finding the reuse of such data to be 'very acceptable.' Of the responses, 4% viewed the sharing of anonymous data to be unacceptable.

The findings suggest a link between respondents' views towards the acceptability of the reuse of data in research and the identifiability of the data, which diverges from the GDPR and DEA's legislative approaches. The GDPR expressly permits the reuse of identifiable information in research. The definition of de-identified data in the DEA may not preclude the reuse of certain forms of personal data. The qualitative data collected during the survey support these findings as some responses linked respondents' acceptance upon sharing only anonymised data. There is an alignment between the survey data and individuals' contributions to the public consultations that were analysed in Part I. There, individuals' views were more focused on using anonymisation and individual control as acceptable responses to developments in the technologies associated with data processing.

Part I found that others' contributions during the consultation process and subsequent practices shaped the legislative responses of the GDPR and DEA. These minimised the impact of these views on the legislative texts. This demonstrates how the 'legitimate expectations of society for the increase of knowledge,'[16] which helps justify a permissive approach to the reuse of data within the GDPR, may not reflect certain views collected during this survey. The identifiability of the data is not the only means of classifying data present in the law. The next section presents the respondents' views on the acceptability of

[16] Recital 113, GDPR.

reusing sensitive data to explore how far these views diverge from the legislative approaches.

6.6 THE ACCEPTABILITY OF REUSING SENSITIVE DATA

This section presents the findings that explore how the sensitivity of the data may influence respondents' views towards the acceptability of reusing data in research. These findings can then be compared to the legislative approaches towards reusing data that are of different sensitivities. The previous section outlined how the GDPR distinguished between different types of data based on their ability to identify individuals. The GDPR also distinguishes between data types based on their sensitivity. Article 9 GDPR creates a unique approach to the processing of 'special categories' of data that are particularly sensitive. The processing of such data is prohibited[17] unless one can demonstrate that the processing falls within one of the exemptions listed in Article 9(2). One of these grounds refers to research purposes,[18] supporting the argument that research is constructed as an activity in the public interest that warrants a permissive governance approach. The DEA does not distinguish between different types of data based on its sensitivity. Therefore, data on an individual's ethnicity or religion could be reused in research where it was not reasonably likely to disclose their identity.

Respondents were asked about the acceptability of reusing a range of administrative data types in research that could fall within the different categories of data found in the law. The list was intended to be illustrative of the

[17] See Article 9(1) GDPR.

[18] See Article 9(2)(j) GDPR.

range of different administrative data, enabling respondents to understand the focus of this thesis better. The list incorporates special categories of data, including information on an individual's religion and ethnicity. Also included are data that are more likely to be anonymous, as defined by Recital 26 GDPR,[19] such as the use of local council facilities.

While the GDPR classes data pertaining to an individual's ethnicity or religion as special categories of data with a more restrictive governance approach, the survey responses suggest most respondents found the reuse of such data acceptable. For ethnicity data, 74.9% found it either very or quite acceptable to be reused in research, and 59% found it acceptable for data on their religion to be reused in research. However, as noted in Part I, research benefits from a presumption of compatibility under Article 5(1)(b) GDPR which allows data to be reused for research purposes where data were initially collected for a different purpose. This suggests that where a purpose initially justifies the use of special categories of data, including religion and ethnicity data, then the presumption of compatibility allows this data to be reused for research purposes without the need for further justification. This suggests there may be some convergence between the legal governance of special categories of data by the GDPR and the survey findings.

The only data types that most respondents found unacceptable to reuse in research were respondents' names (205) and addresses (64.2%.) The GDPR would

[19] 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.

likely treat such information as personal data, thus falling within its provisions' scope. However, it does not treat these data types as particularly sensitive, and they may be reused in research. The approach of the DEA permits the reuse of data where it is unlikely to disclose the identity of an individual. This approach more closely aligns with the survey responses, suggesting that individuals find data types that are likely to disclose their identity unacceptable for reuse in research.

6.7 THE ACCEPTABILITY OF DATA LINKAGE

The ability to link greater amounts of data together was identified in Chapter 1 as a key technological development that facilitates the use of big data analytics on data to find novel insights. Chapter 1 also identified that linking greater amounts of data also prompts new and increased risks to individuals' interests, including their privacy. Nevertheless, the GDPR and DEA remain silent as to the impact of greater amounts of data being linked together. There is no distinction within the activity of processing as to the amount of data that may be linked together, where it is judged necessary to achieve the processing purpose. This approach may be due to the market-making discourses identified in Part I. Part I proposed that the legislative approaches were influenced by dominant discourses that constructed such advances as opportunities that could not be missed as they would significantly benefit both public and private entities.

In contrast, the survey findings suggest that respondents' views towards the acceptability of reusing administrative data in research were influenced by the ability to link certain data. Sixty-eight percent of respondents found it

unacceptable for their name to be linked with other data, and 64.4% found it unacceptable for their addresses to be linked to other data types. These were the most unacceptable types of data to be linked with other types of data, suggesting respondents' views were shaped by the identifiability of the data. Therefore, the more likely a data type would identify them, the findings suggest that more respondents would find the linking of this data to be unacceptable. Respondents' views towards the acceptability of sharing sensitive data and identifiable data suggest that the latter more negatively influences respondents' views towards the acceptability of reusing data in research.

Whilst ethnicity and religion data fall within the GDPR's special categories of data, which are deemed to be particularly sensitive, 66.1% of respondents found it acceptable for the ethnicity data to be linked and 57.7% found it acceptable for their religious beliefs to be linked to other data. Based upon the findings presented above, which suggested a correlation between respondents' views and their ethnicities, respondents' answers were separated based on their ethnicity. Of those identifying from a BAME background, 52.4% found it acceptable for their ethnicity data to be linked compared to 70% of white respondents. Forty six percent of BAME respondents found it acceptable for data on their religious beliefs to be linked to other data compared to 60.8% of white respondents. The following section explores whether the demographic characteristics of respondents might be associated with their views towards the acceptability of reusing data in research. This enables the further analysis of these findings to determine if they are statistically significant.

Aitken and others'[20] research also indicated that older respondents, aged 55 and over, were less comfortable with data being linked for research purposes. The data on respondents' ages were collapsed into two categories, respondents aged 18-50 and 51 and over. This was due to the small number of respondents in some of the categories. Due to the use of categories to collect respondents' ages this was the closest possible replication of Aitken and others' research parameters. Descriptive statistics of the survey items, split based on the respondent's age, were generated, and are presented in Table 7.

[20] Mhairi Aitken and others, 'Public Preferences Regarding Data Linkage for Health Research: A Discrete Choice Experiment' (2018) 3 International Journal of Population Data Science <https://ijpds.org/article/view/429>.

Table 7: Descriptive Statistics Categorised by Age

	Respondents Aged 18-50					Respondents Aged Over 50				
		95% Confidence Interval for Mean					95% Confidence Interval for Mean			
	Mean	Lower Bound	Upper Bound	Median	Range	Mean	Lower Bound	Upper Bound	Median	Range
Your name	3.25	3.01	3.50	4.00	4	2.95	2.57	3.33	3.00	4
Your address	3.42	3.20	3.64	4.00	4	3.11	2.72	3.51	3.00	4
Your date of birth	2.62	2.43	2.82	2.00	4	2.39	2.04	2.74	2.00	4
Your ethnicity	2.10	1.92	2.27	2.00	4	2.23	1.91	2.55	2.00	4
Your income	2.63	2.44	2.82	2.00	4	3.08	2.73	3.44	3.00	4
Your religion	2.43	2.24	2.62	2.00	4	2.51	2.17	2.84	2.00	4
Your education history	2.33	2.15	2.52	2.00	4	2.43	2.09	2.76	2.00	4
Your employment history	2.66	2.46	2.86	2.00	4	2.44	2.11	2.77	2.00	4
Your immigration history	2.66	2.45	2.87	2.00	4	2.57	2.24	2.91	2.00	4
The taxes you have paid	2.63	2.43	2.84	2.00	4	2.82	2.43	3.21	3.00	4
Information about your access to benefits provided by the UK government	2.61	2.41	2.82	2.00	4	2.80	2.42	3.19	2.00	4
Your housing situation e.g., renting or homeowner	2.39	2.20	2.58	2.00	4	2.43	2.09	2.76	2.00	4
The vehicles that you own	2.71	2.51	2.90	2.00	4	2.43	2.08	2.77	2.00	4
Your use of public transport	2.35	2.15	2.54	2.00	4	2.16	1.84	2.49	2.00	4
Your use of public libraries	2.20	2.01	2.38	2.00	4	2.11	1.80	2.43	2.00	4
Your use of public parks	2.34	2.15	2.54	2.00	4	2.20	1.86	2.53	2.00	4

Your use of Council owned gyms, leisure centres, and facilities	2.30	2.11	2.48	2.00	4	2.15	1.83	2.46	2.00	4

The descriptive statistics suggest differences between the respondents' views when separated into these two age groups. The differences in the median suggested that respondents within the younger age category found it more acceptable for their income and tax to be reused, whilst those aged over 50 found it more acceptable for their name and address to be reused in research. The difference in the median for the items concerned with the identifiability of data suggests the respondents aged over 50 found it more acceptable for pseudonymous data and data in the most useful form for research to be reused, compared to respondents in the younger category. The respondents in the younger category found it more acceptable than older respondents for community organisations to reuse administrative data in research. The final survey items concern the linking of different data types. Younger respondents found it more acceptable for their income, tax, and access to benefits data to be linked than older participants. However, older participants found it more acceptable for their name, address, and income to be linked than younger participants. However, this information cannot explain why these differences occur or if they are statistically significant.

A Mann-Whitney test was performed on the survey items to explore if there was a statistically significant association between the respondents' ages and their views towards the acceptability of linking the different types of administrative

data.[21] There were no statistically significant associations between a respondent's age and the acceptability of any of the different data types being linked. The test did reveal some statistically significant differences based on the age of the respondents that are presented in Table 8. Of the items presented in Table 8, younger respondents found it more acceptable for these data types and actors to be involved in secondary uses of data in research except the form of data most useful for research. For this item, respondents in the older category found it more acceptable for this to be reused in research. The survey cannot shed light on why this difference occurs and highlights the need to explore these findings further to better respond to the research questions.

Table 8: Mann-Whitney Test Statistics

	Mann-Whitney U	Z	Asymp. Sig. (2-tailed)
Your income	6543.500	-1.960	0.050
Anonymous data so that I cannot be identified	7001.000	-2.431	0.015
Whatever form of data is most useful for research	6912.500	-2.053	0.040
Central government e.g., HM Revenue & Customs and the Department for Work and Pensions	6684.000	-2.510	0.012

[21] See Annex B for further details.

Local government e.g., County Council and City Councils	6299.500	-3.143	0.002
Charities	6476.000	-2.754	0.006
Community Organisations, which aim to benefit certain groups of people in a community but are not charities	6374.000	-2.914	0.004
Business when working in partnership with the UK government	6855.500	-2.156	0.031

Grouping Variable: Respondents 18-50 and Respondents 50+

Though the Mann-Whitney test suggested that these differences were statistically significant, the effect sizes of these differences were calculated to all be $r<.2$. This suggests that the respondent's age had a small, albeit significant, effect upon their views towards these aspects of reusing administrative data in research. Doing so indicates that other characteristics also influenced the respondents' views towards reusing administrative data in research.

The free-text boxes allowed respondents to provide further details on their views on the acceptability of linking together different data types. Many respondents highlighted concerns over the use of name, date of birth, and address as types of data they found particularly unacceptable. Alongside this were comments that people were wary of data being linked that would render them identifiable. Respondent 59 stated that the 'increased risk of a GDPR breach is too

high compared to the gain of the research with those identifiers.' This suggests that the value placed upon research by the GDPR and DEA, which helped justify a permissive approach to its governance, may diverge from some respondents' valuation of research. Others highlighted the relevancy of the context to their views, with Respondent 14 stating the same data could be helpful in some contexts but 'maybe detrimental if used for other purposes.' The context was also relevant to Respondent 115, who rooted their concern in the possibility of 'excessive or overly nosey data gathering/fishing practices.' These comments suggest a potential lack of trust in the actors involved in the reuse of data in research whereby the aim would not be to contribute to the public interest but rather fulfil other purposes deemed less acceptable.

6.8 The Ability to Meaningfully Speak of 'Societal Expectations'

A key theme throughout this chapter is the diverse range of views towards the acceptability of factors presented in the survey. Such diversity challenges the legitimacy of the law, where it is presumed to be based on a system of socially constructed norms, beliefs, values, and definitions. This potentially undermines the law's ability to draw upon homogenous notions of the 'public interest' and societal expectations' towards increases in knowledge to justify its approach. The findings indicate diversity, ambivalence, and caution towards the actors and purposes relevant to reusing administrative data in social science research.

Analyses explored the potential associations between certain demographics and respondents' responses to explore this diversity further. As the responses were

collected based on a Likert scale, the data was treated as ordinal data and was thus analysed using non-parametric tests.

Previous research in the context of medical research has suggested concerns about the ability of data processing for research purposes to increase stigmatisation and discrimination against the ethnic groups included in the study.[22] Furthermore, the raw data presented in the previous section suggested differences in the acceptability of the different items in the survey. Therefore, further analysis was undertaken to explore whether these differences were statistically significant. Due to the small subgroup sample sizes, ethnicity was collapsed into a binary variable of white and BAME responses. Those who preferred not to provide their ethnicity were excluded from the analysis and treated as missing data for the purposes of this analysis. Descriptive statistics regarding the median of the responses were created to identify differences in the acceptability of the survey items. Table 9 provides the data on the responses and shows differences in respondents' views towards a range of survey items.

[22] J de Vries and others, 'Knowing Who to Trust: Exploring the Role of 'Ethical Metadata' in Mediating Risk of Harm in Collaborative Genomics Research in Africa' (2014) 15 BMC and Medical Ethics 62 and Shona Kalkman and others, 'Patients' and Public Views and Attitudes towards the Sharing of Health Data for Research: A Narrative Review of the Empirical Evidence' [2019] J Med Ethics <http://jme.bmj.com/content/early/2019/11/11/medethics2019105651.abstract>.

Table 9: Descriptive statistics of survey items grouped by White Ethnicity and BAME.

	White Ethnicity Respondents					BAME Respondents				
	Mean	95% Confidence Interval for Mean		Median	Range	Mean	95% Confidence Interval for Mean		Median	Range
		Lower Bound	Upper Bound				Lower Bound	Upper Bound		
Your name	3.21	2.98	3.43	3.00	4	2.68	1.93	3.44	2.00	4
Your address	3.36	3.15	3.57	4.00	4	3.32	2.67	3.96	3.00	4
Your date of birth	2.59	2.40	2.78	2.00	4	2.47	1.80	3.14	2.00	4
Your ethnicity	2.10	1.94	2.27	2.00	4	2.26	1.71	2.82	2.00	4
Your income	2.76	2.57	2.94	3.00	4	2.84	2.19	3.49	2.00	4
Your religion	2.46	2.28	2.64	2.00	4	2.63	2.07	3.19	2.00	4
Your education history	2.36	2.19	2.54	2.00	4	2.53	1.90	3.16	2.00	4
Your employment history	2.59	2.41	2.78	2.00	4	2.68	1.96	3.40	2.00	4
Your immigration history	2.62	2.43	2.81	2.00	4	2.63	1.93	3.34	2.00	4
The taxes you have paid	2.69	2.50	2.89	2.00	4	2.68	1.95	3.42	2.00	4
Information about your access to benefits provided by the UK government	2.71	2.51	2.90	2.00	4	2.79	2.06	3.52	2.00	4
Your housing situation e.g., renting or homeowner	2.35	2.18	2.53	2.00	4	2.68	2.00	3.37	2.00	4

The vehicles that you own	2.58	2.40	2.77	2.00	4	3.21	2.60	3.82	3.00	4
Your use of public transport	2.22	2.04	2.40	2.00	4	3.11	2.44	3.77	3.00	4
Your use of public libraries	2.15	1.98	2.33	2.00	4	2.63	1.97	3.30	2.00	4
Your use of public parks	2.28	2.10	2.46	2.00	4	2.68	1.98	3.39	2.00	4
Your use of Council owned gyms, leisure centres, and facilities	2.22	2.05	2.39	2.00	4	2.84	2.17	3.51	3.00	4
Anonymous data so that I cannot be identified	1.34	1.23	1.44	1.00	4	1.26	1.05	1.48	1.00	1
Identifiable data so researchers can access data that identifies who I am	3.83	3.65	4.01	4.00	4	4.16	3.64	4.67	5.00	3
Data that strongly reduces the possibility of me being identified, but does not make it impossible	2.89	2.71	3.08	3.00	4	2.95	2.28	3.62	3.00	4
Whatever form of data is most useful for research	3.43	3.26	3.59	3.00	4	4.05	3.53	4.57	4.00	3

Central government e.g., HM Revenue & Customs and the Department for Work and Pensions	2.21	2.05	2.37	2.00	4	2.11	1.63	2.58	2.00	4
Local government e.g., County Council and City Councils	2.21	2.06	2.37	2.00	4	2.47	1.91	3.04	2.00	4
Academics at UK universities	2.34	2.17	2.51	2.00	4	3.11	2.37	3.84	3.00	4
Academics at overseas universities	3.05	2.86	3.24	3.00	4	3.79	3.12	4.46	4.00	4
Charities	3.00	2.83	3.18	3.00	4	3.63	3.05	4.22	3.00	3
Community Organisations, which aim to benefit certain groups of people in a community but are not charities	3.06	2.89	3.24	3.00	4	3.68	3.13	4.24	3.00	3
Businesses for their own commercial purposes	4.26	4.11	4.40	5.00	4	4.42	4.02	4.82	5.00	3
Business when working in partnership with the UK government	3.60	3.43	3.77	4.00	4	3.79	3.20	4.38	4.00	3
Your name	3.81	3.61	4.00	4.00	4	3.58	2.82	4.34	4.00	4
Your address	3.57	3.38	3.76	4.00	4	3.84	3.24	4.45	4.00	4
Your date of birth	2.92	2.71	3.12	2.00	4	3.21	2.46	3.96	3.00	4
Your ethnicity	2.26	2.10	2.42	2.00	4	2.84	2.19	3.49	3.00	4

Your income	2.72	2.54	2.91	2.00	4	2.89	2.16	3.63	3.00	4
Your religion	2.48	2.30	2.65	2.00	4	3.05	2.42	3.68	3.00	4
Your education history	2.37	2.20	2.53	2.00	4	2.89	2.20	3.59	2.00	4
Your employment history	2.58	2.41	2.76	2.00	4	3.05	2.34	3.76	2.00	4
Your immigration history	2.66	2.47	2.84	2.00	4	3.16	2.42	3.90	3.00	4
The taxes you have paid	2.83	2.64	3.03	2.00	4	3.00	2.32	3.68	3.00	4
Information about your access to benefits provided by the UK government	2.86	2.66	3.05	2.00	4	3.32	2.61	4.02	3.00	4
Your housing situation e.g., renting or homeowner	2.46	2.28	2.64	2.00	4	3.00	2.34	3.66	3.00	4
The vehicles that you own	2.69	2.51	2.88	2.00	4	3.21	2.54	3.88	3.00	4
Your use of public transport	2.26	2.09	2.42	2.00	4	3.05	2.36	3.74	3.00	4
Your use of public libraries	2.23	2.07	2.40	2.00	4	2.74	2.10	3.38	2.00	4
To improve my life	1.83	1.69	1.96	2.00	4	1.84	1.55	2.13	2.00	2
To improve the lives of others	1.77	1.64	1.91	2.00	4	1.79	1.53	2.05	2.00	2
To help the government see where it can improve public services	1.78	1.65	1.91	2.00	4	1.79	1.45	2.13	2.00	3

To help the government see where it could save money by making public services more effective or efficient	2.13	1.98	2.28	2.00	4	2.16	1.70	2.62	2.00	3
To increase our knowledge of the challenges facing society	1.92	1.78	2.05	2.00	4	2.05	1.61	2.49	2.00	3
For any purpose the government would like	4.08	3.94	4.23	4.00	4	4.16	3.70	4.62	4.00	3

A Mann-Whitney test was performed on the data to explore whether the differences shown in Table 9 were statistically significant. The analysis demonstrated that some of these responses were statistically significant, as demonstrated in Table 10.

Table 10: Statistically Significant Results from the Mann-Whitney Test

	Mann-Whitney U	Z	Asymp. Sig. (2-tailed)
Your use of public transport	2516.000	-2.321	0.020
Your use of Council owned gyms, leisure centres, and facilities	2421.000	-2.324	0.020
Academics at UK universities	2400.500	-2.752	0.006
Academics at overseas universities	2589.500	-2.229	0.026
Charities	2671.500	-2.027	0.043

Your housing situation e.g., renting or homeowner	2690.000	-1.991	0.046
Your use of public transport	2662.500	-2.079	0.038

Grouping Variable: Ethnicity Grouped into White and BAME Respondents

Whilst these differences were found to be statistically significant, the effect size of these differences was small to medium (r<-.2)[23] for all the survey items, suggesting that other factors also influenced respondents' views towards the acceptability of reusing administrative data in research. All other differences between the medians identified in Table 9 were found to be non-significant through the Mann-Whitney test.

Robinson and others[24] study found that older people tended to be less trusting of commercial organisations. The survey data was analysed to determine if there was an association between respondents' ages and their views towards the acceptability of commercial organisations reusing administrative data in research. A Kruskal-Wallis test was performed on the data after responses that declined to provide their age were coded as missing data. Due to lower numbers of respondents aged over 60, the categories were collapsed into three categories that encompassed respondents aged 18-30, 31-50, and the over 50's. The results suggest that the acceptability of businesses reusing administrative data in research, either

[23] Jacob Cohen, 'Statistical Power Analysis and Research Results' (1973) 10 American Educational Research Journal 225.

[24] Gillian Robinson and others, 'Public Attitudes to Data Sharing in Northern Ireland: Findings from the 2015 Northern Ireland Life and Times Survey' (*Ulster University* February 2018) <https://pure.ulster.ac.uk/en/publications/public-attitudes-to-data-sharing-in-northern-ireland-findings-fro>.

for their commercial purposes or when working in partnership with the UK government, was not significantly affected by a respondent's age. For businesses for their commercial purposes, $H(2)=3.366$, $p=.186>.05$ and for businesses working in partnership with the UK government, $H(2)=4.663$, $p=.097>.05$. However, the study by Robinson and others focussed on trust rather than acceptability. Therefore, one explanation for these differences is that other factors may influence the respondents' views towards the acceptability of these purposes despite a low level of trust, but this cannot be proven nor disproven by the data.

Robinson and others'[25] study also indicated a correlation between gender and trust in organisations. A Mann-Whitney test was performed on the data, with gender as the grouping variable to determine if these findings were reflected in the data. Due to the very low numbers of respondents who selected options that were neither male nor female, these responses were excluded from the analysis. The survey findings indicated that women found it more acceptable for overseas academics, charities, community organisations, and businesses (both for profit and in partnership with the UK government) to reuse their data. The analysis shows that female respondents' acceptability of the different actors did not differ significantly from male respondents' views in this survey. For central government departments $U=7156$, $z=--1.201$, *ns*, $r=-0.07$, for local government $U=7195$, $z=-1.128$, *ns,* $r=-0.07$, for academics at UK universities $U=7142$, $z=-1.186$, *ns,* $r=-0.07$, academics at overseas universities $U=7755.5$, $z=-.154$, *ns,* $r=-.01$, charities $U=7438$, $z=-.683$, *ns,* $r=-0.04$, community organisations $U=7324$, $z=-.872$, *ns,* $r=-$

[25] Ibid.

0,05, businesses for their commercial purposes U=6960, z=-1.608, *ns,* r=-0.9, and for businesses working in partnership with the UK government U=6920, z=-1.558, *ns,* r=-0.09. The results, therefore, suggest the difference between genders was not significant and represented very low-sized effects. The differences between these results and Robinson and others' findings[26] may be attributable to the difference between the respondents finding an actor's reuse of their data acceptable, yet still having a low level of trust in that actor.

These findings cast doubt on the ability of the DEA to justify the reuse of data in research as based on the 'public interest' and the GDPR's reliance on expectations. However, not all the diversity of views collected in this research could be attributed to the respondents' demographic characteristics. Furthermore, the effect size of statistically significant differences based on different demographics was small. These findings suggest that a range of factors contribute to the diversity of responses. These underlying reasons are difficult to investigate through quantitative methods, a point explored in further detail below, which reflects upon the limitations of the survey and how these findings inform the next stage of the research.

6.9 DISCUSSION

This section situates the research in the broader literature that explores individuals' views towards the reuse of data in research to see how the findings compare. This enables a richer understanding of the contribution of this research towards better understanding individuals' views in this context.

[26] Ibid.

6.9.1 Comparing the Findings to the Existing Literature

This chapter's key findings shed light on respondents' views towards their understanding of research as an activity in the public interest. As noted in Part I, the construction of research as an activity in the public interest justified the permissive governance approaches of the GDPR and DEA as acceptable. The analysis of the survey responses suggests there are some divergences from this construction of research. Respondents' responses suggest that in some areas, they find a less permissive approach to be acceptable. This section will explore the existing literature in this area to compare the findings to existing research. Potential explanations for differences between the findings and previous empirical work are provided to situate this research in the existing literature better.

The strong preference for reusing anonymous data in research suggests a divergence between the legislative instruments and respondents' views towards defining research as an activity in the public interest. This finding resonates with previous research that suggests anonymisation is an expected standard where administrative data are reused in research. Waind[27] notes that much of the existing literature supports the finding that participants are more accepting of their data being reused where it had been anonymised. For example, research on public attitudes to data linkage indicated that respondents generally felt that 'anonymisation' was a key part of the data linkage process and felt it was of great importance that this part happened at the earliest point possible.'[28] The desire

[27] Elizabeth Waind, 'Trust, Security and Public Interest: Striking the Balance' (2020) 5 International Journal of Population Data Science.

[28] NatCen Social Research, 'Public Attitudes to Data Linkage' (2018) <https://natcen.ac.uk/media/1595557/public-attitudes-to-data-linkage.pdf>.

may be linked to the findings of Cameron and others where participants expressed concerns that the 'ideological and political factors that influence what data is collected by government, which in turn influences the way data is interpreted, and the decisions government makes.'[29] Fears of how the government uses data, and the potential influence of factors that respondents disagree with, may have prompted them to define acceptable reuses of administrative data in research as using anonymous data only.

The findings showed that some respondents from minority backgrounds found certain potential aspects of reusing administrative data in research more unacceptable than respondents who identified as white. This suggests that certain respondents had a different understanding of what constitutes research as an activity in the public interest than others which resonates with previous empirical studies. Cameron and others'[30] research suggested that participants from certain backgrounds may be more fearful that such research may be used to 'pigeonhole' certain sections of society or reduce further support given to certain people and/or areas. The heightened unacceptability towards the reuses of administrative data in research may be due to concerns that such data may contribute to discriminatory treatment that pigeonholes or neglects certain sections of society. Aitken and others' empirical research support this contention as it found the potential for the labelling of individuals or groups and 'for such labelling to result in stigma or discriminatory treatment'[31] to be a key concern in the context of

[29] Cameron and others (n 6), 23.

[30] Ibid.

[31] Aitken and others (n 1), 93.

sharing and linking data. Neither the GDPR nor the DEA expressly prohibits reusing data in research in the way that those from a BAME minority were more likely to find to be unacceptable. This suggests that the legislative approaches do not align with the views of certain sections of society towards the acceptability of reusing sensitive categories of data. However, the analysis suggested that there was no statistically significant difference towards reusing administrative data in research based on a respondent's ethnicity in many instances. Furthermore, where statistically significant differences were found, the analysis suggested the effect size of these differences were small.

In contrast to the GDPR's approach, the research purpose appeared to be a relevant factor in respondents' constructions of research as an activity that justified the reuse of their data. The purpose appears more relevant to the approach of the DEA, where researchers must demonstrate that their work is in the public interest before they may be granted access to administrative data. Nevertheless, there appear to be divergences between the dominant discourses driving the DEA's legislative process, which shaped a definition of research in the public interest as biased towards quantitative research that provided tangible benefits aligning with the wants of public authorities and respondents' views. The survey responses demonstrated that most respondents found the reuse of administrative data in research acceptable when it benefited them and others in society. A majority, however, also found it acceptable for administrative data to be reused where it did not entail a tangible beneficial outcome for individuals but did increase knowledge about societal challenges. In doing so, it does not support previous empirical work in this area, which suggested participants' support was

premised upon research 'undertaken to deliver an outcome of social value.'[32] The same study found that '[b]lue skies or theory-led academic research continued to be valued far less than research perceived as socially beneficial (i.e. research that has the potential for clear policy or practice impact).'[33] Furthermore, in Cameron and others' study, participants criticised research that lacked a tangible point.[34] The findings in this chapter suggest that the survey respondents had a broader definition of an outcome of 'social value' than those participating in Cameron and others' research.

Nevertheless, the mixed responses towards the acceptability of reusing whatever form of administrative data that is most useful for research purposes suggests the social value arising from reusing administrative data in research does not necessarily translate to finding this research acceptable. This may be due to respondents finding research that contributes to existing knowledge is acceptable but remain concerned about how research would be used in practice. In doing so, their concern translated to the uses of whatever data would be most useful for research purposes being treated with ambivalence. This ambivalence supports a distinction between this use of data and the reuse of health data in medical research. Previous research suggested suggests that individuals define expert access to their health data as an unquestionable benefit.[35] These findings intimate

[32] Cameron and others (n 6).

[33] Ibid.

[34] Ibid.

[35] Wellcome Trust, 'Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data' (*Wellcome Library,* July 2013) <https://wellcomelibrary.org/item/b20997358#?c=0&m=0&s=0&cv=0>.

this cannot be extrapolated to all research using other types of data. This aligns with Cameron and others'[36] study, where some participants linked reusing administrative data to implementing controversial policies, such as the bedroom tax.[37] These links may further challenge the perceived acceptability of the reuse of such data. Other studies have indicated some are sceptical about whether research utilising reused data would lead to any government or policy changes.[38] The findings of this study potentially suggest that whilst respondents theoretically supported a broad understanding of research, their concerns about what would happen in practice prompted a more restrictive approach to be deemed more acceptable.

The findings suggest a tension between the broad definition of research as an activity in the public interest and the responses that found the reuse of data most helpful in research to be an unacceptable use of data. One potential reason for this may be the trust that respondents had in those involved in the research. Therefore, whilst they had a broad definition of research as an activity in the public interest, they may have been concerned about the research in practice. This resonates with previous findings by Davidson and others,[39] which suggested

[36] Cameron and others (n 6).

[37] Chaminda Jayanetti, '50,000 More Homes Hit by Britain's Bedroom Tax, as Scientists Warn of COVID-19 Risk' *openDemocracy* (February 23, 2021) <https://www.opendemocracy.net/en/opendemocracyuk/50000-more-homes-hit-bedroom-tax-scientists-warn-covid-19-risk/>.

[38] NatCen Social Research (n 28).

[39] Sara Davidson and others, 'Public Acceptability of Data Sharing between the Public, Private and Third Sectors for Research Purposes' (2013) <https://www.gov.scot/publications/public-acceptability-data-sharing-between-public-private-third-sectors-research-purposes/pages/13/>.

respondents' views towards the acceptability of reusing data in research was influenced by the actors involved in the data processing. The findings from this survey suggest that government actors were found to be the most acceptable actors, resonating with Davidson and others' study. 40 Their research found 'there was an assumption that public bodies had more stringent data protection and security procedures in place than other types of organisation – or at least were more accountable to the public.'[41] This echoes previous findings that suggested participants had certain expectations that 'by having access to more data and research, public bodies will be better placed to identify and meet the support needs of particular segments of the population.'[42] This suggests an alignment between respondents' views and the DEA's approach that creates a permissive gateway where public authorities can choose to allow researchers to access previously collected administrative data.

Nevertheless, other studies have identified 'complex or ambivalent relationships of trust' towards actors not in the private sector.[43] Davidson and McLean identified that many participants in their study 'generally trusted public bodies more than commercial organisations…a significant minority of participants held considerably more sceptical views.'[44] This may explain the decrease in

[40] Ibid.

[41] Ibid.

[42] Ibid.

[43] Aitken and others (n 1).

[44] Sara Davidson and Christopher McLean, 'Public Acceptability of Cross-Sectoral Data Linkage' (2012) <https://www.gov.scot/publications/public-acceptability-cross-sectoral-data-linkage/>.

support where governmental actors partnered with businesses. Previous studies have found many are concerned about the role of commercial organisations where their relevance to research is 'acknowledged,' but concerns remain.[45] Aitken and others' study echoed the concern about profits, and participants also voiced a low level of trust in the private sector but, still, 'such concerns did not necessarily mean outright opposition to private sector involvement in research.'[46]

The exceptional approach towards reusing data in research within the GDPR and the DEA's exclusive gateway for accredited researchers suggests these instruments construct researchers' identities as relevant to the governance of administrative data, including its disclosure and reuse. In this context, the provisions of the DEA and the GDPR suggest their construction of a 'researcher' justifies disclosure and reuse for research purposes. The findings in this chapter suggest there was no universal support for the acceptability of researchers reusing data in research. Instead, there was significant diversity in the views towards the reuse of data by overseas academic institutions. There is very little in the GDPR regarding the actors involved in reusing data for research purposes. However, researchers are impliedly constructed positively due to the positive construction of research in the GDPR and the reference to the research's high ethical standards. This broad-brush approach does not align with the empirical findings. The DEA requires researchers to fulfil several criteria to become accredited, which includes

[45] Mhairi Aitken, Sarah Cunningham-Burley and Claudia Pagliari, 'Moving from Trust to Trustworthiness: Experiences of Public Engagement in the Scottish Health Informatics Programme' (2016) 43 Science and Public Policy 713.

[46] Aitken and others (n 1).

expertise in quantitative research and to be acting in the public interest. It is not clear whether respondents' views on the acceptability of researchers at universities was driven by respondents' views towards researchers' abilities to act in the public interest.

The responses to the acceptability of linking data suggest a divergence between respondents' views and the legislative approaches, which do not explicitly acknowledge the risks of linking data to individuals' privacy interests. This is a further example of the survey responses, suggesting that respondents felt a more restrictive approach to the governance of reusing data in research was acceptable. This more restrictive construction of what is acceptable is supported by the findings that indicated most respondents felt reusing anonymous data was the most acceptable approach, their definition of research as an activity in the public interest as narrower than suggested by the GDPR and DEA. Furthermore, the responses suggest that respondents' definitions of researchers, as an acceptable actor to reuse administrative data, were narrower than that present in the GDPR and potentially the DEA. Studies have indicated that public support is at least partially conditioned on data being used for 'legitimate purposes.'[47] These findings suggest that some may find what is lawful to be illegitimate. The findings presented in this chapter support a more cautious approach to the reuse of administrative data in social science research than is suggested in the permissive approach identified in Part I, undermining legislative justifications based on societal support.

[47] Ibid.

These findings, therefore, raise questions as to how far the law governing the reuse of data may be deemed 'legitimate' in the sense of permitting actions deemed acceptable or desirable by reference to a socially constructed set of norms, beliefs, and values. This point is critical as previous research has suggested that perceptions of legitimacy are a condition of individuals' support for the reuse of data in research. What remains unclear, however, is how this legitimacy is to be defined. This definitional issue has been noted in previous research.[48] The difference between these findings and previous empirical work suggests that some respondents may hold different views on social science research's value and 'legitimacy', irrespective of its capacity for a 'real world' impact. Previous research has suggested that public support and acceptability of research is linked to it being 'in the public interest.'[49] The findings from this study provide some support for the notion that individuals' expectations regarding research may not require the inclusion of tangible outcomes as an increase in knowledge may be sufficient to justify the research as acceptable. This suggests a greater alignment with the GDPR's approach, which constructed research as an activity in the public interest, irrespective of its outcomes. These findings challenge the dominant discourses that shaped the legislative approach of the DEA, which appears to have a narrower definition of research that is acceptable to reuse administrative data. Chapter 4 demonstrated that research that justified the reuse of data within the DEA is quantitatively focussed, and the criterion of the 'public interest' is potentially defined in alignment with the desires of public authorities, including

[48] Aitken and others (n 1).

[49] Ibid.

their financial needs. There was support for the reuse of administrative data in research that sought to improve knowledge rather than demonstrate a tangible outcome.

However, public support for research that improves knowledge should be treated with caution as other studies have acknowledged the nebulous nature of the concept of the 'public interest,' engendering a 'varied and shifting' interpretation. [50] There were particular concerns that certain actors could interpret the public interest to support their own ends or for commercial gain, which did not align with participants' definition of the public interest. [51] Such findings may help explain why the involvement of businesses was felt to be particularly unacceptable, even when working in partnership with the UK government. Davidson and McClean's empirical research suggested that participants felt the notion that the 'public interest' could be used to justify 'virtually any use of linked data.'[52] This indicates there are risks of the public becoming disillusioned with data reuse in research where it is felt to justify certain activities that may not align with their view of what is acceptable or desirable.

The findings thus demonstrate that an understanding of legitimacy underpinned by a single system of socially constructed norms, beliefs, values, and definitions may be difficult to achieve in this context. Whilst a true societal consensus on a topic may be very difficult to achieve, some of the responses suggest

[50] Davidson and McLean (n 44), 18.

[51] Ibid.

[52] Ibid.

that the reuse of administrative data in research prompts a particularly heterogeneous response. The potential for a disparate impact may undermine the acceptability of a majoritarian justification for the governance of reusing administrative data in research.

6.9.2 Taking the Findings Forward: Acknowledging Limitations and the Need for Further Research

This chapter provided several findings regarding respondents' views towards the acceptability of different aspects of reusing administrative data in research. This enabled the identification of several divergences from the GDPR and DEA's respective approaches towards this data reuse. The survey findings shed some light upon the factors that shape acceptability, but further work must be undertaken to explore how they shape expectations and the role of other factors that are difficult to investigate in a survey. This section discusses the limitations of this research and the further research suggested as necessary by the findings.

A survey can provide insights into respondents' views but is limited in how far it explains why respondents feel a certain way. The free-text boxes within the survey allowed respondents to provide further information, which shed light on some of the reasoning used by respondents to justify their views. Nevertheless, they were insufficient alone to provide sufficient data on respondents' underlying views and beliefs that shaped their responses. Qualitative research is more suited to exploring and constructing the layperson understandings of certain terms, which can feed into how they justify their views. This is critical as the understanding of legitimacy in this thesis is based on what is deemed acceptable

and the system of values, beliefs, norms, and definitions that construct what is understood to be acceptable.

In particular, the findings suggested the importance of the public interest and the legitimacy of the reuses of administrative data in research as critical to securing respondents' support. The survey, however, did not provide the necessary freedom for respondents to provide their definitions of these terms and how these terms are understood in the context of reusing administrative data in research. Further research exploring the perceived benefits of social science research using administrative data would be useful. This could shed light on individuals' distinctions and how this informs their views towards this activity's acceptability. Qualitative research better enables respondents to provide definitions in their own words that are shaped by their personal experiences, providing insights not just into what people think but why they think a certain way. In doing so, such research can build upon the quantitative findings to explore what people think and the views, beliefs, norms, and definitions that underpin their construction of acceptable reuses of administrative data in research.

Alongside the boundaries of the 'public interest' in rendering the reuse of administrative data in research, further research exploring the factors that shape individuals' views and fears could be beneficial. A richer understanding of these elements could shed further light on the system of norms, values, and beliefs that individuals use to determine whether the reuse of administrative data in research is a desirable or acceptable activity. It could provide further insights into what prompts the diverse range of views collected by the survey to understand better what makes the reuse of administrative data so contentious.

Furthermore, the free-text responses highlighted a further limitation of the survey as the qualitative data suggested the interaction between the different factors shaped respondents' views, yet the survey presented them to respondents in isolation. The multi-faceted nature of individuals' assessments of the acceptability of the reuse of data in research aligns with the findings by Aitken and others.[53] In their study, whilst there were 'clear patterns in which actors were generally considered to be more or less trustworthy…this pattern did not straightforwardly translate into support for research conducted by healthcare professional or academics.' Instead, support depended 'on a range of factors.' This suggests a need to investigate these findings further through a method that allows individuals to discuss the interaction of different factors and the aspects of a context. These findings suggest that trust is a factor that shapes individuals' views towards the acceptability of reusing administrative data in research. The survey explored respondents' views towards the acceptability of their involvement in the reuse of administrative data. However, it did not explicitly explore the relevance of trust to respondents' views, nor the role trust can play in rendering certain reuses of data in research more or less acceptable to individuals. The complex or ambivalent relationships of trust' towards actors identified in previous research suggests this would be better suited to a research method that allowed participants more freedom to discuss and debate their views. Doing so would respond to the second research question explored by this thesis that asks why divergences between what is legally permissible and what is found to be acceptable in the empirical work may occur.

[53] Aitken and others (n 45).

6.10 CONCLUSION

This chapter presented the qualitative findings of the research that suggested several divergences between what respondents found to be acceptable reuses of administrative data in research and the approaches of the GDPR and DEA. Several qualitative findings suggested that respondents found a more restrictive approach towards the governance of the reuse of data in research was constructed as acceptable than the approach reflected in the law. The discussion drew upon previous research to explore potential reasons for these divergences, highlighting the importance of legitimacy and the different understandings of the public interest in the reuse of data in research. There were, however, areas where the findings did not align with previous research. Furthermore, the survey responses cannot explore why such divergences between the legal approach and respondents' views may occur. This suggests that further research is required to investigate the findings further. These findings will inform the next research stage by suggesting areas where respondents' views diverged from the legislative approaches. The following two chapters present the qualitative research, which enables a richer understanding of individuals' views towards the acceptability of the reuse of administrative data in research and how it compares to the approaches of the GDPR and DEA.

7 QUALITATIVE FINDINGS I: EXPLORING THE PROBLEMATISATIONS ARISING FROM THE FOCUS GROUP DISCUSSIONS

7.1 INTRODUCTION

The findings from Part I and the previous chapter identified possible divergences between the legal approach and societal expectations towards reusing administrative data in research. Part I analysed the approaches of the GDPR and the DEA, the processes that shaped the texts, and their effects. Part II has, thus far, highlighted the potential role of societal views in securing the law's legitimacy and has presented the quantitative findings. This suggests that there are possible areas of divergence between the law and societal approaches towards what is acceptable in the context of reusing administrative data in research. Chapters 7 and 8 present the qualitative findings to explore these points further, which provides a richer understanding of participants' views and their underlying justifications. This chapter builds on the previous chapter's findings to shed further light on their views through participants' own words and personal experiences. The findings help elucidate participants' problematisations of the governance of reusing administrative data in research. This enables a deeper exploration of how these diverge from the problematisations identified in Part I and informs potential responses to the identified divergences and their potential causes.

Five online focus groups explored participants' views on the appropriateness of reusing administrative data in research and the views, beliefs,

norms, and definitions underpinning these views.[1] The data analysis applied Bacchi's framework [2] to the empirical data to explore participants' problematisations of the governance of the reuse of administrative data in social science research. The diversity in participants' responses makes it only possible to explore prominent, rather than dominant, discourses. Nevertheless, they still provide useful data on participants' problematisations and how they diverge from the problematisations justifying the GDPR and DEA's respective approaches.

This chapter's findings shed light on participants' problematisations of the reuse of administrative data. In contrast to the GDPR and DEA, the lack of a need for consensus generated multiple problematisations, and participants remained in disagreement about what was appropriate. Nevertheless, the findings suggest key themes that shaped participants' problematisations, providing insights into the divergences between participants and the GDPR and DEA and the underlying reasons for them. Unlike the practices that shaped the legislative approaches, participants' responses suggested they were more cautious and ambivalent towards this use of data. As a result, participants were less likely to construct the reuse of data in research in such a positive light and questioned whether it was in the public interest. The chapter explores the findings regarding the assumptions that underpin these more cautious problematisations, analysing the impact of participants' mistrust of certain actors and the concept of data ownership. Finally, findings on external events and economic interests as further factors that

[1] See Annex A and Annex C for further details.

[2] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

influenced problematisations that differ from those identified in Part I are discussed.

The conclusion reflects upon the findings to introduce the next chapter, which draws upon Bacchi's framework to explore the effects of these findings. These effects include some participants finding a less permissive approach towards reusing administrative data in research than present in the current law being appropriate and a diversity in views. Part III examines the impact of clashing discourses that lead to different views on what is appropriate in this context to consider how the law can respond to such diversity.

7.2 JUSTIFYING THE USE OF ONLINE FOCUS GROUPS

Focus groups are an apt method for exploring divergences between the legal approach and societal expectations in what is deemed appropriate in this context and why these divergences may occur. The understanding of legitimacy considered throughout this thesis suggests that the question of why these divergences occur maps onto divergences between the underlying system of norms, beliefs, values, and definitions that inform those approaches. The exploration of what people felt to be appropriate, and the norms, beliefs, values, and definitions that inform those assumptions, requires a methodology that enables participants to articulate their views in their terms and draw upon their own experiences. This aligns with the strengths of qualitatively-oriented research methods.[3]

[3] J Kitzinger, 'Qualitative Research: Introducing Focus Groups' (1995) 311 BMJ 299.

The emphasis on intra-group interactions and the minimisation of the active participation of the facilitator within the focus group methodology responded to the needs of this thesis. By fostering debate and discussion between participants, further sharing of personal experiences may be encouraged. This can push the discussion further by preventing participants from focussing on the expert status of the facilitator.[4] Through interactive debate and discussion, focus groups are a flexible means to explore opinions and ideas that would be impossible to replicate in a one-to-one interview.[5] Furthermore, during the session, one can collect data on how participants' views and attitudes shift, develop, and respond to ideas proposed by others.[6] The lockdown restrictions necessitated the use of online focus groups. Further information regarding the practical aspects of the planning, recruitment, questioning route and analysis can be found in Annex A and Annex C contains the full questioning route. This chapter focuses on the findings that shed light on participants' problematisations of the governance of reusing administrative data in research.

7.2.1 The Analytical Approach

NVivo[7] was used to transcribe the focus groups, including participants' verbal and non-verbal responses to the discussions. The transcriptions included pauses,

[4] Ibid.

[5] EA Whitley, N Kanellopoulou and J Kaye, 'Consent and Research Governance in Biobanks: Evidence from Focus Groups with Medical Researchers' (2012) 15 Public Health Genomics 232.

[6] Ibid.

[7] QSR International, 'Qualitative Data Analysis Software | NVivo' (2021) <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>.

repeated words, and other examples of speech disfluency[8] to help identify where participants were more hesitant or formed their responses as they spoke. It was important to present the findings authentically both the words used and how they were articulated provide data relevant to the research questions. The data analysis drew upon Bacchi's framework[9] and the understanding of legitimacy[10] referenced throughout the thesis. The use of Bacchi's framework enabled a comparison between how the legislative instruments and focus group participants problematised the reuse of administrative data in research, and the effects that cascaded from their respective problematisations. This best facilitated the identification of divergences and the reasons why these occurred. Doing so enabled a reflection upon the legitimacy of the GDPR and DEA's approaches towards reusing administrative data in research. The findings suggest that the legitimacy of the law, premised upon assumptions about the appropriateness of an action based on a socially constructed system of norms, beliefs, views, and definitions, might be challenged in this context. The focus groups helped expose participants' diverse views, undermining the idea of a single system of beliefs concerning appropriate reuses of administrative data in research, which permits a discussion of prominent, rather than dominant, discourses.

While the analytical approach towards the qualitative data is a novel use of Bacchi's framework, it enables a richer reflection upon the divergences between

[8] Wikipedia Contributors, 'Speech Disfluency' (2019) <https://en.wikipedia.org/wiki/Speech_disfluency>.

[9] Bacchi (n 2).

[10] Mark Suchman, 'Managing Legitimacy: Strategic and Institutional Approaches' (1995) 20 Academy of Management 571.

participants' views and the legislative instruments. The framework's focus on discourses remains relevant to the analysis of focus groups to explore the debates, discussions, non-verbal displays of emotion, and how participants' views interacted and evolved. The findings chime with Bacchi's point that discourses create competing representations that shape conceptions, perceptions, and preferences.[11] Though no consensus was achieved during the focus groups, as an agreement is not an aim, and no single discourse truly dominated the discussions, it was still possible to identify prominent discourses and their underlying assumptions. The notion of prominent discourses suggest that alternatives were not rendered entirely appropriate or unthinkable but were heavily influential in shaping participants views during the focus groups.

7.3 PARTICIPANTS' PROBLEMATISATION OF THE REUSE OF ADMINISTRATIVE DATA IN RESEARCH

7.3.1 Opportunities Tempered by Risks

The dominant discourses that shaped the legal problematisations of the reuse of administrative data in research identified in Part I emphasised the opportunities presented by technological advances in this field. Such discourses influenced a permissive legislative response as the appropriate 'solution' to this use of data, including the discourses highlighting the innovative nature of these advances and the opportunities they presented in unlocking new insights into societal challenges. However, such discourses did not dominate the focus groups. Instead, the prominent discourses of participants highlighted the potential risks and disadvantages arising from such research, though some recognition of the

[11] Bacchi (n 2).

opportunities was still present. The risks concerned negative implications for the participants themselves, but also the wider society. Nevertheless, this still diverges from the problematisation that justified the legislative responses, which enabled a permissive approach to data reuse in research.

One prominent discourse that highlighted the perceived risk of this reuse of data focussed on the potential for discriminatory outcomes. This was particularly prevalent in participants from a minority ethnic background. Many such participants felt very suspicious due to their beliefs as to its potential for negative discriminatory impacts. P4FG3 (Participant 4, Focus Group 3) stated that

> *my main concern is because I am from an ethnic origin which is not British, that's my problem because…there can be biases… anything, anywhere, you fill any form they ask you, you do any survey they ask you what's your ethnic origin and I always think to myself, why do they want to know my ethnic origin, you know, what's the reason? Because I'm not British…*

Such views were echoed in different focus groups, providing further evidence of the prominence of this discourse. The emphasis on discrimination, as part of some participants' problematisation of the governance of reusing data in research, is highlighted by P3FG2's comment that

> *I'm very easily profiled, I wear a hijab and I'm you know, hah, I feel like specifically for me, I don't want anyone profiling me, like, just as easy as you can do just because of the colour of my skin, or because of how I dress or and and the life decisions I make so I'm very wary about these things and it doesn't sit well with me*

These findings contrast with the quantitative findings, which identified differences in participants' views towards the acceptability of reusing ethnicity

data based on their ethnicity. However, the analysis suggested these were not statistically significant differences. The focus groups refine and expand upon the findings through comments that reinforce the reception of such data reuses not as opportunities but as sources of concern. These findings suggest both the presence of a divergence but also why such a divergence has arisen. P7FG1 expressed concern about the risk of 'false correlations based on a small sample of the population so...they could end up disproportionately affecting ethnic minorities.' These findings align with previous empirical research by Cameron and others[12] that found those from certain backgrounds were particularly concerned about being 'pigeonholed' by research that fed into policy decisions. These backgrounds were not just concerned with participants ethnicities and concerned a range of factors that may have led them to fear being profiled or pigeonholed. These factors may explain why those from a BAME background raised these concerns during the focus groups, whilst the quantitative findings did not suggest a statistically significant difference in the views of respondents based on their ethnicities. The differences between the problematisations within the prominent discourses of the focus groups and those identified during the legislative processes of the GDPR and DEA give insights into why participants may not find the permissive approach of the law to be an appropriate solution.

The prominence of risk-focused discourses arose not just in the context of racial discrimination but also the broader potential for discrimination and

[12] Daniel Cameron, Sarah Pope and Michael Clemence, 'Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes' (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/>.

increasing existing inequalities. These discourses may help explain the qualitative findings that found no statistically significant difference in respondents' views based on their ethnicity. This discourse stands in stark contrast to the construction of research in the GDPR and DEA as an activity particularly capable of responding to societal challenges, especially where innovative analytical techniques were used. In contrast, P4FG1 felt there was 'a possibility of [research] being targeted…there is a possibility of ermmmm the research not really benefitting other people.' P3FG2 echoed this notion of potentially being 'targeted' as a risk fostered by developments in technology. P3FG2 stated that 'you're the target, you're the information they are gaining, you are the profit for them and that, that is what concerns me.' Previous research[13] has suggested a link between citizens' construction of this research activity as being in the public interest and the perception that the benefits outweighed the potential risks. The findings suggest that the prominence of the risk-based discourses meant that participants did not necessarily construct reusing administrative data in research as an activity in the public interest. This indicates an aspect of reusing data in research where some participants' views diverged from the GDPR and DEA's construction of what is appropriate. This provides evidence of differences in the underlying beliefs regarding the opportunities presented by the innovations used by researchers upon administrative data and potentially the value attributed to this activity.

[13] Elizabeth Waind, 'Trust, Security and Public Interest: Striking the Balance' (2020) 5 International Journal of Population Data Science.

Nevertheless, other discourses did not problematise the reuse of administrative data in research as a solely negative development. Some participants recognised both the opportunities and the risks. This ambivalent attitude is present in the comment made by P1FG1 that 'overall, yes, I think [the UK government] have got good intentions but…you know…ummmm there's always a risk of a hidden agenda, isn't there?' P2FG3 echoed this, feeling that 'in theory they have the best intentions…but, in reality, it'll probably be an annoyance.' These concerns support the findings made by Cameron and others[14] that the perceived social value of research is critical to public support. The ambivalence and suspicion present within the focus group discussions suggest certain participants may not perceive such research as having sufficient social value to be deemed an appropriate reuse of their data. This provides evidence of a divergence between the problematisation of this activity, whereby participants view the inclusion of economic interests as a negative factor in their construction of the appropriateness of reusing data in research. This link between the perceived riskiness of this activity and the potential for economic interests to be served by the data processing activity suggests that participants did not necessarily share the market framing approach of the GDPR and the DEA identified in Part I.

Only one participant's views resonated with the dominant discourses identified in Part I. P3FG4's support appeared to be rooted in their lived experiences as a self-employed businessman. P3FG4 echoed the discourses identified in Part I, emphasising the benefits of research in helping them find the best employees for their business, demonstrating their green credentials to secure

[14] Cameron and others (n 12).

contracts, and providing opportunites to be involved in pilot schemes that could further benefit their business. In their view, the risks discussed by others did not bother them, and the reuse of data in research enabled 'Britain to lead the way' to become more environmentally self-reliant. Nevertheless, P3FG4 was in the minority, with several participants within that focus group acknowledging but disagreeing with this discourse. Furthermore, P3FG4 acknowledged that whilst they were very open about their data, 'it's different really for each person.' When probed further about their views, their views developed slightly away from the focus on opportunities. The focus on the opportunities seemed to be underpinned by an attitude of trying not to let government uses of their data 'bother' them as 'if you let it bother you too much you can get quite caught up in it. I think its link….it is a worry though, definitely.' This shift suggests that the discursive effect of the focus group rendered this view less appropriate and more 'unthinkable' compared to the prominent discourses.

Several participants contributed to a discourse that was more cautiously supportive of this use of data. P5FG1 acknowledged that

> *I don't mind the research at all but when its controlled…when I cannot control how much I give and how much is used or what I give, it's not as good…but, yeah, I support it because it can make our lives better and change some policies or bring new policies to light and yeah, it's important, it's important to do it the right way*

When probed further as to the 'right way,' P5FG1 confirmed this entailed the use of consent. The desire for an 'opt-in' mechanism suggests a more granular construction of research whereby only certain examples are appropriate. This

diverges from the presumption of compatibility present in the GDPR[15] and the approach of the DEA. P6FG1 echoed this by recognising the opportunities for the research to be for 'a good cause' as insufficient to make the reuse of identifiable data appropriate. P5FG1 was opposed to reuses of data in research that could facilitate a 'personal trace back.'

This section provides evidence of divergences between the prominent discourses that shaped the problematisations of focus group participants and the legislative instruments regarding the opportunities arising from reuses of data in research. The prominent focus group discourses highlighted the potential risks fostered by this use of data, raising concerns regarding 'targeting' and racial discrimination alongside the possible opportunities. This led to a more cautious and tempered assessment of the reuse of data in research than present in the legislative processes of the GDPR and DEA, which may shed light on the different 'solutions' explored in the next chapter. The findings question the law's legitimacy in this context due to the beliefs and values expressed by participants that were not so present during the legislative processes, which may lead to divergences in what is deemed appropriate and proper.

7.3.2 The Dark Side of Innovations

The prominent discourses of the focus groups also highlighted a 'dark side' to the innovations constructed so positively in the problematisations underpinning the GDPR and DEA. Again, participants focussed on the potential harms, diverging from the problematisations that justified the permissive approaches of the

[15] Article 5(1)(b) GDPR

legislative instruments. The focus group problematisation did not construct the developments in data processing technologies as something to capitalise upon but rather a source of concern. For example, P1FG3 stated that

> *technology was supposed to set us free and it has done anything but! It has made our lives far more complicated, we're now overloaded with information, most of which we do not need, we are targeted left, right and centre*

P5FG3's comments suggested a construction of 'innovation' as a source of concern rather than an opportunity. P5FG3 felt that 'all of our information is in there as well and I just feel it makes us easier to trace and tracked and errrm…..identified by all these different things that are connected with.' The prominent discourses referenced technological innovations to problematise innovations as a source of concern that undermined participants' control over their lives. P4FG4 felt that 'you just don't know, even when you're walking down the street, you don't know what camera's following you and what kind of data is being gathered.' This developed into a discussion where P3FG4 felt the developments meant 'it's a bit like Big Brother's watching over you but that's the way it is, really.' While P3FG4 found such developments beneficial, aligning with the prominent discourses shaping the GDPR and DEA, the other participants were unconvinced about the benefits of technology in modern society. This broad concern prompted participants to problematise the specific example of the reuse of administrative data in research as a potentially unwelcome development.

A discourse regarding the positives of innovation can be found in the 2015 Conservative Party Manifesto,[16] which shaped the DEA's approach. Within that text, the novel interactions between academia and industry were constructed as positive developments producing beneficial outcomes, including financial profit for universities. This problematisation constructs financial gain as a positive outcome of research and constructs interactions between researchers and commercial entities as a source of positive opportunities. Chapter 4 suggested how these discourses influenced the DEA's approach towards the governance of reusing administrative data in research. However, the prominent discourses within the focus group suggest that participants' problematisations diverged from this through their construction of such developments as inappropriate and undesirable. For P5FG3 they fostered an

> *incestuous thing going on between companies and government where they can match this information to build up a profile of you and I just think 'how the hell did we get there from, I don't know [laughs]' I wish we could go back but we've opened a Pandora's box, it seems*

A prominent discourse within the focus groups linked the appropriateness of a data processing activity to the identity of the actors involved. This led to distinctions not necessarily present in the law, and responses to these distinctions that diverged from the GDPR and DEA's approaches. A prominent discourse concerned the undesirability of commercial actors being involved in reusing data in research. Rather than constructing their involvement as an opportunity, the responses suggest many felt their inclusion would undermine the potential public interest in such research. For example, P1FG3 felt that commercial entities were

[16] The Conservative Party, 'The Conservative Party Manifesto 2015' (2015) <https://issuu.com/conservativeparty/docs/ge_manifesto_low_res_bdecb3a47a0faf>.

'beholden to their shareholders' and unconcerned with acting in the 'public interest.' This supports the qualitative findings where most participants found the inclusion of commercial actors to be unacceptable. The focus groups provide further insights regarding the underlying reasons, suggesting beliefs regarding the motivations of such actors as a source of concern. The findings suggest an area of divergence between the problematisations present in the focus groups and underpinning the approaches of the GDPR and DEA.

The discourses regarding the appropriateness of academic researchers and charities were more supportive than discourses regarding other actors. Participants were more likely to find their involvement to be appropriate, but this support was neither universal nor unqualified. The DEA constructed relationships between commercial actors and universities as a positive development but not all participants shared this approach. For P1FG5, the commercialisation of academic institutions prompted a negative reaction which triggered a desire for a stricter approach towards universities gaining access to such data. They felt that

> *probably 25 years ago, universities were academic bodies and I would probably have had a lot less of an issue with this sort thing because I think it would have been for very genuine reasons er I kinda feel now that universities are operating as businesses… universities are primarily making money to pay, you know big salaries to their senior staff and I kind of start to question whether a university has got the same errrrmmm kind of cloak around it that it would have had not that many years ago.*

Through such comments, academic researchers were problematised so as not to justify the permissive approach present in the GDPR. The DEA limits access to administrative data to researchers who can demonstrate that their research is

in the public interest.[17] However, whether the 'public interest' is constructed as mutually exclusive from any commercial interests remains unclear. The discourse within the Conservative Party Manifesto, which welcomed the amalgamation of these domains, suggests it may diverge from the prominent discourses within the focus groups. The focus groups found the inclusion of commercial actors to be an inappropriate, rather than a positive, development.

Such findings indicate divergences in the responses towards the innovative nature of modern data processing practices. In contrast to the emphasis upon the beneficial outcomes and the opportunities fostered by alliances between academic and commercial entities present in the problematisations underpinning the law, the prominent discourses within the focus groups characterise such developments in a more negative light. This suggests that participants may find the permissive approach towards the governance of reusing data, which was partially justified based on these discourses, inappropriate.

7.4 THE ASSUMPTIONS UNDERLYING THE PARTICIPANTS' PROBLEMATISATIONS

Building upon the previous section's findings, this section will explore the underlying assumptions and presumptions that justified participants' views. It examines how these foster discourses within the focus groups diverge from the discourses that shaped the legislative approaches. This sheds light on the underlying differences in the norms, values, beliefs, and definitions that determine what is appropriate in this context. This section presents findings on

[17] As required in the Code of Practice, established by s70 Digital Economy Act 2017.

assumptions regarding the untrustworthiness of certain actors, assumptions regarding the value of research in this context, and a presumption of individuals owning their data.

7.4.1 The Trustworthiness of Different Actors

Previous research highlighted 'trust' as a key condition for public support,[18] but the legislative instruments are silent on the role and relevance of trust in determining the legality of reusing administrative data in research. The findings support the previous research[19] in this area. The development of the provocations introduced new actors that highlighted their presence as a factor that shaped participants' views towards the appropriateness of reusing administrative data. Such distinctions suggest divergences between the approaches of the participants and the GDPR and DEA. Part I identified a broad-brush approach in the GDPR towards the definition of 'researchers' that could potentially permit research by commercial actors to benefit from the permissive approach towards the reuse of administrative data in research in the GDPR. The DEA narrows its permissive approach towards reusing administrative data to research in the 'public interest.' However, as identified in the previous section, this does not necessarily preclude the involvement of commercial entities. The Accreditation Criteria that complement the DEA is premised upon a researcher's expertise and willingness to

[18] Cameron and others (n 12); and Mhairi Aitken and others, 'Public Preferences Regarding Data Linkage for Health Research: A Discrete Choice Experiment' (2018) 3 International Journal of Population Data Science <https://ijpds.org/article/view/429>.

[19] Mhairi Aitken and others (n 18); and 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 <https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-016-0153-x>.

undertake certain training rather than their affiliation with a certain institution.[20] These distinctions do not reflect those present in the prominent discourses of the focus groups, suggesting an area of divergence.

7.4.1.1 University Researchers

For many participants, university-based researchers were assumed to be one of the more trustworthy actors in this context, and many found their involvement in the reuse of data in research as appropriate. P3FG2 stated that 'I trust universities more than probably...say, if it was a different type of researcher for example the government or something else.' Underpinning this discourse was the perception that university-based researchers operated in accordance with high ethical standards. P8FG3 stated that 'the fact that its university-based increases my trust because I know universities have to go through, like, ethics boards and all that stuff so, for me, that would increase the trust.' For P2FG2, this specifically justified their differing perceptions towards the actors involved in reusing administrative data in research as

> *maybe the government will omit data that is gunna benefit their conclusions more in a very...biased way, potentially than a university researcher might do because they're bound to this ethical review that they have to use as a process.*

The trust fostered by the assumptions towards university researchers' pursuit of high ethical standards aligns with some of the discourses present in the GDPR, which referenced the 'recognised ethical standards for research.'[21] Further

[20] UK Statistics Authority, 'Data Sharing Code of Practice and Accreditation Criteria: Research' (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567074/Research_Code_and_Accreditation_Criteria__1___2_.pdf>.

[21] Recital 33 GDPR

supporting the discourse of university researchers as trustworthy institutions were assumptions as to the neutrality fostered by academic research. This is suggested in P8FG5's comments where

> *there's no hidden agenda maybe, there's no business and profit here, it's more for, you know, not genuine but err productive and constructive reasons i.e., research to better things in erm education and all that that side of things.*

However, as shown in the previous section, not all shared this view of academic researchers within universities. P2FG5 echoed this sentiment through their worry about 'academic motives and stuff because society and universities have changed a lot over the past century so who knows what the next trend will be.' Furthermore, several participants' comments suggested trust alone was insufficient to gain their support for their data to be reused for research as permitted by the current law. Some still felt that, alongside this ethical process, there was a need for consent. P5FG3 stated that the

> *problem isn't the university having my information, it's just the consent, it's the fact that the information is just being picked up randomly and we don't have any knowledge of that and it's not that I don't trust the university, it's just the process that I don't trust here...'*

Furthermore, certain aspects of this data processing activity seemed to undermine some participants' trust in university researchers. The concept of research happening without consent seemed to undermine the presumption of trust some had in university researchers. For example, P5FG1 stated

> *I know how researchers are bound to protect participants' privacy and it would just be really difficult for me to…to realise that…all my contribution was just violated errrm in a sense and it was just given out like that*

The qualitative findings support previous research suggesting that a higher level of trust does not necessarily translate into automatic support for such actors to reuse data for research purposes.[22] This suggests that the presumptions and assumptions underpinning participants' views on what is acceptable in this context are more granular than the legal approach analysed in Part I. The findings also suggest that participants' construction of what is appropriate in this context may depend upon more factors than considered relevant in the provisions of the GDPR and DEA. Aitken and others' research suggested that the data sharing processes, transparency, and accountability mechanisms were also highly relevant to how supportive people were towards research.[23] The provocations facilitated the findings on participants' views towards the different actors involved in reusing administrative data in research. The provocations were developed by introducing different actors to gauge how participants' views would develop that facilitated an exploration of the assumptions underpinning these changes. The following section explores how this shed light on participants' views towards commercial entities.

7.4.1.2 Commercial Entities as 'Beholden to Shareholders'

The importance of the actor's identity in shaping participants' views towards the appropriateness of reusing data in research is reflected in P1FG4's comments. To them, the development of the provocation to include a commercial entity

> *does change the context because university researcher will be just looking at the data for what it is, looking at the answers, analysing them whereas a business is probably...gunna have some confirmation*

[22] Aitken and others (n 19), 720.

[23] Ibid.

bias, they're probably gunna want to skew the data potentially to align with their own...erm ethos and try and maximise their money.

This level of mistrust towards commercial actors expressed in this statement represents a prominent discourse present during the focus groups. Many rooted their mistrust in their perceptions as to the priorities of commercial entities. P3FG2 stated, 'I trust private sectors less than I do the government actually, they have more of a vested interest in my eyes than the government do to basically sell you a product.' The presumption that the pursuit of profit undermined the ability of commercial actors to operate in a way that participants would trust to be in their interests was prominent throughout the focus groups. P1FG3 said of commercial actors that 'their first and only purpose in life is to make money. They're a commercial entity, that is all they're interested in doing.' As a result, P1FG3 felt that 'it is the shareholders that are the people that will benefit at the end of the day, not necessarily the people being researched.' Other focus groups echoed this discourse, emphasising participants' views towards commercial entities as an inappropriate actor in the context of reusing administrative data in research. P1FG4 felt commercial entities would 'use this data to achieve their own goals and earn more money and it wouldn't be for the benefit of the people, it would somehow be used in a way that suited them financially.'

Despite the prominence of the presumption that commercial entities were less trustworthy than other actors in this context, this did not mean all participants felt they should have no access to administrative data for research purposes. Some supported their access to findings, as opposed to identifiable data. P4FG5 stated 'that's really positive…actually a change can still come out of the

data.' Such findings chime with research undertaken by Tully and others [24] whereby the benefits arising from such research were the main driver of participants' views. This enabled commercial actors to be viewed as appropriate where they contributed public benefits rather than just private profits. [25] However, the prominent discourses within the focus groups did not reflect this. Instead, many constructed the involvement of such actors to be inappropriate, suggesting the benefits commercial actors could produce was insufficient to garner participants' support. Unlike Tully and others' study, this research did not provide participants with access to partial and impartial actors as this research aimed to explore participants' views rather than informed citizens.[26] This difference may shed light on the responses to the identified divergences, which Part III explores. The focus here is upon the problematisation underpinning the responses. The findings in this section suggest participants problematised actors differently to the legislative instruments, driven by low feelings of trust and differing definitions of the benefits that could arise from their involvement.

7.4.1.3 The Government's Role

The legal gateway created by s64 DEA allows public authorities to determine whether administrative data should be shared with researchers. The presumption that public authorities are the best-placed entity to determine which actors it is appropriate to share data with underpins this approach. This suggests a certain

[24] Mary P Tully and others, 'Commercial Use of Health Data—a Public 'Trial' by Citizens' Jury' (2019) 3 Learning Health Systems.

[25] Ibid.

[26] Ibid.

level of trust in public authorities to make such decisions based on society's best interests. However, the findings suggest this was not a prominent discourse within the focus groups. This may suggest a divergence in the problematisation of the governance of the reuse of administrative data that may prompt participants to prefer a different solution than the permissive gateway within the DEA.

However, this study's qualitative and quantitative findings differ in the responses regarding the acceptability of the government reusing administrative data in research. The prominent discourses within the focus groups emphasised the low levels of trust towards the government to reuse data in a way that met their definitions of the public interest. In contrast, 72.6% of the survey participants found it either very or quite acceptable for central government to reuse administrative data for research purposes, and 69.0% of participants found it either very or quite acceptable for local government bodies to do the same. Those who participated in the focus groups were asked to complete the same survey to ensure all participants had the same baseline level of knowledge to contribute to the focus group discussions. The focus group participants were less supportive, with 58.3% finding it either very or quite acceptable for central government to reuse administrative data in research. Furthermore, 48.3% of participants finding it either quite or very acceptable for the local government to do the same. It is not possible to provide a definitive explanation for the differences between the survey responses, nor the differences between the initial survey findings and the focus group data. It is important to highlight the number of focus group participants (n=36) renders the findings from the survey that they completed less robust than this thesis' main quantitative piece of research. Furthermore, these findings may

reflect the impact of several external effects that occurred between the two stages of the empirical research. This is explored in further detail below.

A further potential reason that may partially explain the differences between the quantitative and the qualitative findings is that the discussion-based format of focus groups allows participants to develop their views considering others' experiences and opinions. This occurred in FG5, where P3FG5 responded to the first provocation by saying

> *trouble is there's a huge trust issue these days because...there's been so many scandals with people's information being sold on or being used for different purposes and you wouldn't hope it would happen when it was the government but even the government have been...you know...there's been issues with...lobbying and commercial deals, I think there's a lot, a lot, a big lack of trust with the way data's being misused...*

P4FG5 stated in response that

> *when I was reading it, I was thinking that...it all sounded fine and…that...it all looked okay and I hadn't even considered what [P3FG5] said, it didn't even enter my mind but now that he's mentioned it...actually...maybe it is something to think about errrm, yeah, I was just, yeah, it just didn't enter my mind at all but I guess like [P3FG5] said there have been scandals that maybe we need to think about and whether that would affect this study.*

P4FG5's response suggested P3FG5's comments prompted a shift in their views, aligning more with the prominent risk-based discourses discussed above. Furthermore, for some participants this mistrust was rooted in their beliefs about the government's interests in this area. P3FG1 stated that they did not 'always believe the government has the best motives.' Other focus groups echoed this sentiment. P1FG2 did not 'altogether trust the government to be acting in the

national interest' and felt there was 'a big chunk of data that the government has and you haven't a clue what it is really – probably best not knowing.'

Low levels of trust in the government to act appropriately in the context of reusing administrative data in research was a prominent discourse throughout the focus groups. In contrast to the prominent discourses that focussed on researchers discussed above, participants seemed to construct the government as an actor more prone to using research in a more biased way. P3FG1 felt that 'if the government has an idea what they want to implement, it's very easy for them to select what they sort of focus on or what they highlight and sort of ignore other things.' This chimes with comments made by P8FG5, who felt

> *most of the governments in the work have hidden agendas and when the findings come out of the research and it does not suit their agenda, the government have the upper hand [so] it can decide to just forego the findings.*

The comments suggested participants were less trusting of the government, and the government's involvement shaped their response to the governance of reusing administrative data in research. The dominance of these discourses means that many participants' problematisation led to a solution that would not give government entities the same 'gatekeeping' powers as present in the DEA.

Nevertheless, some recognised the potential for the government to act with the best intentions, and the mistrust expressed did not entirely prevent the recognition of the benefits that could arise. Such ambivalence and cautious support in the face of these low levels of trust chimes with Aitken and others' research that emphasised the importance of trust for securing support yet

acknowledged the relationships of trust that operate in this context are 'nuanced and complex.'[27]

This chapter's findings suggest that the differing levels of trust in various actors shaped participants' problematisation of the governance of reusing administrative data in research. The actors that participants particularly mistrusted shaped the problematisation in two ways. Firstly, where participants had low levels of trust in an actor that may reuse the data in research, many felt such actors should only have access to anonymous data or, for some, no access to any data. This was particularly prominent when participants discussed commercial actors accessing administrative data where the pursuit of profit and the importance of shareholders influenced several participants' problematisation of the governance of administrative data differently from the current law. Secondly, the low levels of trust that many participants expressed towards the government meant their problematisation of the governance of the reuse of administrative data did not necessarily align with the DEA's creation of a permissive legal gateway that granted public authorities the power to grant access. These findings suggest that the norms and beliefs that influenced participants' problematisations of the governance of reusing data in research were less influential in shaping the problematisations identified in Part I. Consequently, there may be differences in what is permitted by the law and what participants construct as appropriate. Doing so raises questions about the legitimacy of the law in this context. Trust in the actors involved in reusing administrative data in research was not the only area where trust appeared

[27] Aitken and others (n 19).

pertinent to participants' views. The following section will present findings that explore participants' views towards the trustworthiness of research.

7.4.2 Regarding the Trustworthiness of Research

The discourses that underpinned the GDPR and DEA suggested that research, particularly where it capitalised upon innovative data-driven technologies, contributed to society by enabling new insights into societal problems. In doing so, such research was constructed as an activity in the public interest. Therefore, the problematisations that shaped the legal response to the reuse of data in research justified a permissive approach to enable the benefits of research.

The GDPR presented a very positive view of research, focussing on its possible contributions to resolving seemingly intractable societal issues[28] and constructing academic outputs as a facet of the freedom of expression.[29] The DEA's approach limited the benefits of its governance approach to accredited research in the public interest. Underpinning this approach were assumptions about the value of research and trust in its ability to enable better responses to societal issues. Nevertheless, as Chapter 4 discussed, the practices that shaped the DEA constructed research as an activity worthy of greater access to administrative data that was more quantitatively oriented. The qualitative findings within this research suggest differences in the trust individuals had in research that reused administrative data.

[28] See, for example, Recital 157 GDPR.

[29] Article 85, GDPR.

In contrast to the GDPR, where research is presented as a standalone activity with little focus on the actors involved, participants did not separate the activity of research from the actors that undertook it. Therefore, due to the low levels of trust in the actors, as discussed in the previous section, many participants also had low levels of trust in the ability of research to provide public benefits. As noted previously, existing empirical research on the reuse of data in research has suggested that the belief that such research is in the public interest and can provide societal benefits is critical to securing public support.[30] Certain comments made by participants suggested the role played by the government contributed to their low levels of trust in research being able to provide public benefits. One interaction highlighted this view whereby P7FG2 stated that this use of data 'sounds like it's in the public interest' to which P1FG7 responded 'until the government gets its hands on it and spins it…' No participant opposed this view, and several expressed support for this view through non-verbal means, including nodding and smiling. The mistrust in research due to mistrust in actors involved in reusing data in research is echoed P1FG3's comment that local authorities have 'either ignored the research that's been done or used it to pedal what they were planning to do in the first place.'

The government was not the only actor whose involvement led participants to suggest they had low levels of trust in the research that could take place. P4FG4 felt researchers had their 'own implicit biases and prejudices, and that's the way everyone is but I feel like…it's just one of the ways we allow biases to trickle down.' These findings shed light on the participants' different beliefs, which shaped their

[30] Waind (n 13).

understanding of what may be appropriate in this context. These beliefs shaped assumptions towards the impact of the involvement of certain actors and their definition of research as potentially not in the public interest. Doing so reveals differences in participants' problematisations of the governance of reusing data in research and the problematisations that justified the legal responses analysed in Part I. As a result, the findings in this thesis suggest a divergence between the legal construction of the activity of research and participants' views, leading several participants to question the appropriateness of research reusing administrative data.

The qualitative findings suggest the elision between the activity of research and the ability of research to solve problems, which shaped the approaches of the GDPR and DEA, was not prominent within the focus groups. This suggests that when participants problematised the governance of the reuse of administrative data, they did not assume that research would automatically contribute to society in the way present in the discourses that shaped the legislative responses. Instead, many participants expressed a more ambivalent attitude towards the research outcomes, simultaneously acknowledging its potential to benefit and harm individuals. For P1FG4, there was mistrust in research due to their negative definition of its outcomes. P1FG4 felt that 'the linking of [data] implies that it's not just the data, like its conclusions being drawn, based off of two different things but together so there might be assumptions made about people.' P7FG2 was cautious in their support for research with no elision between the activity and the possibility of positive outcomes.

> *I think for me it comes down to what impact, err, how's that data gunna be used and what's the impact on my life? If it's gunna be in these policies and trying to make the area I live in better, then it's a good thing, but say if its gunna err go to the extent that it's gunna profile me…*

Others' problematisations of the governance of reusing administrative data in research echoed this feeling and was used to justify a more granular approach towards the activity of research. This suggests a different definition of research underpinned participants' views that did not construct it so homogenously than the construction of research suggested by the GDPR and DEA. P2FG2 hoped

> *that the greater good for society would outweigh those risks and therefore it would be worthwhile, but I also agree with the point that it's difficult to give a definitive answer when you don't know the aims of each research project.*

Such comments support the existing empirical research[31] that emphasised the importance of the perceived public benefits arising from such research in securing citizens' support. The prominent discourses suggest the definition of research diverged from that used within the GDPR and DEA in terms of its ability to foster public benefits. The different beliefs about the value of research may shed light on how far participants' views diverged from the permissive approach present in the DEA and GDPR, which was partially justified by discourses that constructed research as a beneficial activity. In Chapter 4, the analysis of the practices that shaped the provisions of the DEA and the accompanying Accreditation Criteria suggested a bias towards quantitatively oriented research being constructed as an appropriate reuse of data. However, the prominent discourses within focus groups diverged from this approach. P7FG1 drew upon

[31] Tully (n 24).

their experiences to disagree with this approach. P7FG1's comments suggest the importance of purpose as a factor shaping their construction of the appropriateness of reusing data in research.

> *I've seen examples of how data has been used incorrectly by the government, so research purpose would take precedence because this is very reductive data, it's not qualitative so it's not gunna include everything so there's a question of how much it would be useful at face value.*

Others' comments suggested a similar low trust in quantitative research. When discussing the potential for research to occur without an individual's consent, P1FG3 felt it meant people were

> *being looked at almost, in a harsh way, like some sort of laboratory animal that's being put under the microscope…it dehumanises everything because you're looking at numbers, at data, and we all know the expression about 'lies, damned lies, and statistics.*

For P2FG2, the reuse of data in research could enable the 'context' to be removed, which appeared to undermine their trust in research to respond to societal challenges

> *as when there's so much context missing about that and the data could be kind of taken, bits of the data could be taken by the researcher and manipulated…[used] in a based way rather than kinda seeing data in the picture of the person as a whole.*

These findings suggest divergences in the problematisation of the governance of the reuse of administrative data in research present in the focus groups and legal provisions analysed in Part I. In contrast to the problematisations analysed in Part I, many participants appeared to have low levels of trust in research being able to provide the types of benefits that justified the permissive legislative approaches. This suggests that the value attributed to research as an activity in the public interest by participants is lower than that

underlying the GDPR and DEA's approaches. The divergence in defining research as an activity in the public interest may have implications for the legitimacy of the legal governance of the reuse of data in research. A comparison between the findings of Part I of the thesis and the qualitative research suggests this construction justifies a permissive governance approach that many participants seemed to feel was inappropriate and risked undesirable consequences.

7.4.3 Regarding the Concept of Data Ownership

A final assumption that shaped how participants problematised the governance of reusing data was the concept of 'data ownership.' The language of ownership is not present within either the GDPR or DEA, which suggests a divergence in the problematisation of data governance. A prominent discourse concerned 'data ownership' as a concept that justified a high level of individual control over individuals' administrative data. The concept of 'ownership' continued to influence some participants' views towards secondary uses of administrative data.

Whilst some attributed an importance to the concept of ownership as applying to their data, several recognised the potential for negative consequences to arise.

> *I'd like having the autonomy over whether I can withhold or give my data and I'd like to know what it's gunna be used for because I think, on the one hand, I think withholding my data could be a form of protest against that research if I don't want that kind of research to be carried out but, on the other hand, I wonder if I'm going to…further…effect the research by not providing the data because then they're not gunna get an accurate reflection of the population. [P7FG1]*

Furthermore, not all participants problematised the governance of the reuse of administrative data in a way that required such stringent levels of

individual control. For several participants, their reasoning suggested this level of control was too burdensome and would not engage with it. P8FG2 felt that when the government collected data, 'it's not our data anymore, is what I feel, it's out of our hands, what happens with it really isn't our choice.' Initially, P8FG2 felt that where people had consented to their data being collected, they had relinquished their data ownership. When probed as to how that would apply in the context of the reuse of administrative data in research, where there is no legal requirement to gain an individual's consent, their views remained unchanged. They reasserted that this data had, at some point, been 'given' to the government. Several of the participants within this focus group disagreed because their problematisation of reusing data in research required some individual input. For P4FG2, this was rooted in the belief that 'the information has come from us…it is our information...it's useful because we've given it so we should have some control over it.'

In problematising the governance of the reuse of data in research as an activity that must respect an individual's data ownership, there is a significant divergence from the problematisations identified in Part I. The presence of this concept in the focus groups sheds light on some of the norms that participants drew upon when determining what is appropriate in this context. Whilst not a universally held view within the focus groups, its presence raises questions about the legitimacy of the GDPR and the DEA in this context where individual input is near non-existent. Nevertheless, the diverse range of views undermines the ability of the GDPR and DEA to reflect a societally constructed system of norms as the

qualitative findings suggest a single system is lacking. Part III explores the implications of this in the context of reusing data in research.

7.5 THE FACTORS PROMPTING THE PARTICIPANTS' PROBLEMATISATIONS

The final section will present the qualitative findings that shed light on how the factors shaped participants' problematisations of the governance of the reuse of administrative data. This section will focus on the role of external events and the response towards the market-framing approach of the GDPR and DEA.

7.5.1 External Events

Throughout the focus groups, participants referenced high-profile scandals involving those within the government that, whilst not directly connected to the reuse of data in research, still shaped their problematisations. Previous research has highlighted the importance of 'mass media in the origin of public trust.'[32] The findings suggest that external events influenced the problematisation by undermining the trust participants had in certain actors. For example, P2FG5 mentioned that

> *I read something about [the government] being presented with lots of findings and case studies about cladding on buildings after the whole Grenfell Towers incident and that's still ongoing so I think sometimes governments don't always have errrrm I guess, like, the best interests of people at the top of their list.*

P5FG3's responses suggested their trust in the government had been undermined by the Greensill lobbying scandal[33] as, to them, it undermined the

[32] Tully (n 24).

[33] Danica Kirka, 'UK Lobbying Scandal Snares Ex-PM Cameron' (*AP NEWS,* April 25, 2021)

government's claims to neutrality. Instead of looking after the public's best interests, the scandal has made it seem to P5FG3 that the government instead preferred to their friends and that 'I've seen enough to know that it's a bit of a dodgy area.' P3FG5 attributed their feelings to

> *huge trust issue these days because…there's been so many scandals with people's information being sold on or being used for different purposes and you would hope it wouldn't happen when it was the government but even the government have been…you know…there's been issues with…lobbying and commercial deals.*

Furthermore, to some participants, the traditional mechanisms of accountability were felt to no longer be capable of helping to secure sufficient trust in government to justify the reuse of data in research as appropriate. P2FG2 stated

> *I think if we think about coronavirus and bodies and kinda like the bodies we thought were external from the government and sort of finding out they're more connected to the government then maybe that undermined the trust in general in the 'independent' bodies and how independent they are, so I think has compromised trust for me and the independent bodies are maybe related to the government.*

The blurring between the public and private sectors also contributed to this mistrust, with P1FG3 stating that

> *we've seen with the PPE scandal, with Covid, just how much you can trust the government and their decisions to hand out contracts to private companies so I, I, I…would be extremely unhappy [with data being used this way.*

This echoes previous research by Davidson and others,[34] which found some participants alluded to a blurring between the public and private sectors that they regarded as having negative implications.

The findings suggest that external events influenced a problematisation of the governance of the reuse of data in research that did not necessarily justify the permissive approach identified in Part I. The external events that helped shape participants' values and beliefs emphasised the low trust participants had in the actors involved. This led many to question the public benefits that could arise from this use of data as they felt the benefits would accrue to a select number of actors with existing links to the government. The links between low levels of trust and the perceived value of this use of data align with previous empirical studies that have suggested these factors are important in securing societal support.[35] Previous empirical research has also highlighted the importance of trustworthiness, as distinct from trust, in this context. Aitken and others' study[36] suggests that the extent to which an actor is deemed trustworthy strongly influences participants' responses. The events referred to in this section would undermine the perceived trustworthiness of different actors. In challenging the idea that the research made possible by technological advances would entail public benefits, external events seemed to have sparked a divergence in the value

[34] Sara Davidson and others, 'Public Acceptability of Data Sharing between the Public, Private and Third Sectors for Research Purposes' (2013) <https://www.gov.scot/publications/public-acceptability-data-sharing-between-public-private-third-sectors-research-purposes/pages/13/>, 63.

[35] Waind (n 13).

[36] Mhairi Aitken, Sarah Cunningham-Burley and Claudia Pagliari, 'Moving from Trust to Trustworthiness: Experiences of Public Engagement in the Scottish Health Informatics Programme' (2016) 43 Science and Public Policy 713.

participants and the law attribute to research. This seems to lead to different problematisations about how to govern reuses of data, with participants seeming to disagree with the permissive approach present in the GDPR and DEA.

External events also influenced participants' problematisations by raising questions as to the potential opportunities arising from research. Several participants referred to external events to justify their views as based on their previous experiences. Their experiences led them to conclude that nothing would change due to this use of data, thus not providing the level of public benefits necessary to secure their support for this use of data. P1FG3 felt they would have 'serious problems' with the reuse of administrative data in research because of

> *[h]ow local authorities have responded to similar types of surveys that have been carried out locally…they've either ignored the research that's been done or used it to pedal what they were planning to do in the first place.*

Within another focus group, similar sentiments were echoed, based on participants' views towards the National Census, which had taken place a couple of weeks before the focus groups. The findings suggest this event influenced certain participants' views, such as those expressed by P3FG2, who stated that

> *I'm thinking I haven't seen any changes from the census being used other than loads of facts and figures in loads of reports to say 'Ooooh and this is how many people do this' and…it's just to give it a show basically. I don't know what they use the Census for or what difference it makes.*

Their views on the Census led to a problematisation of the governance of reusing data that highlighted a need to see a 'local difference' and accord some level of individual control. This suggests a problematisation that diverges from those identified in Part I. In contrast to the approaches of the GDPR and DEA,

participants do not appear to focus on the opportunities presented by the advances in big data analytics that may be used upon administrative data sets to provide new insights into solving societal challenges. Instead, external events have prompted a sense of suspicion about these developments in many focus group participants and a questioning of the potential for such research to improve people's lives.

The qualitative findings indicate that the external events that participants used to engage with the provocations led them to question the public benefits that could arise from this use of data. The low levels of trust, sparked by participants' previous experiences, combined with the lack of tangible benefits arising from current research, meant many seemed to dispute the benefits arising from research that shaped the legislative responses. The external events may not have been rooted in previous experiences of the reuse of administrative data in social science research. Therefore, they provide useful insight into the types of experiences participants rely upon when justifying their views. It suggests what shaped the values and beliefs that participants draw upon when discussing and justifying their views and sheds light on how these values and beliefs diverge from what underpinned the legislative responses.

7.5.2 The Market Framing

The discussion of economic interests within focus groups suggested a different problematisation than the market framing approach of the GDPR and DEA that resulted in a permissive approach to the reuse of administrative data in research.

In contrast to the dominant discourses underpinning the DEA and GDPR, there was little regarding the 'burdens' that could be borne by commercial entities where they processed data. Furthermore, there was nothing in the focus group discussions about how the protection of data led participants to play a more active role as online consumers. Instead, the problematisation of the governance of the reuse of administrative data in research in the focus groups suggested a strong opposition towards supporting economic interests in this context. For many participants, the pursuit of economic interests and profits was incompatible with their views on the appropriate reuses of administrative data in research.

The discussions suggested that many felt economic interests were not aligned with the reuses of data that participants would find appropriate because it led to actors being 'beholden to shareholders' rather than acting in the public interest. P6FG5 felt 'I'm sure there is potential for a company to misuse [findings and data] and not in a…for the good of the people, particularly in the local community…It may be for the gain of them in an economic way.' In discussing the provocation where research with administrative data was undertaken to respond to employment issues, P6FG5 stressed, 'I know the research is focussing on inequalities, but I still feel like companies could misuse it.' Therefore, the findings align with previous research[37] that suggested the pursuit of profit is regarded by many as inappropriate in this context. In doing so, this shed some light on the norms, values, and beliefs that shaped participants' understanding of the public interest in this context. Whilst a unified definition of the 'public interest' has not

[37] Tully (n 24) and Davidson and others (n 34).

been identified, the findings suggest that many viewed the pursuit of financial profit as incompatible, leading to such reuses of data to be deemed improper.

Furthermore, in contrast to the 2015 Conservative Manifesto that initially set out the government's plans in this context, participants did not particularly welcome an overlap between industry and academia. P1FG5 distinguished between research that 'is something done speculatively by the researcher to be able to sell into government' and 'whether it's being commissioned by the government.' For them, the pursuit of profit and the idea that 'universities are operating as businesses' led to a problematisation that resulted in a less permissive approach to reusing data in research than present in the law. P5FG5 echoed this, feeling that 'universities, now, they just want more money.' Similarly, the idea that universities may have economic interests led several participants to problematise the governance of the reuse of administrative data requiring stringent controls to protect individuals' interests and the public interest.

These findings suggest that some of the norms associated with academic research have evolved in response to the increasing commercialisation of the academic domain. In contrast to the problematisations that justified the permissive approaches in the GDPR and DEA, this development was not regarded so positively by participants. This sheds light on an area where differences in the norms and values underlying the approaches towards governing the reuse of data diverge as the responses to potential economic interests fostered were different. Whilst the problematisation underlying the legislative provisions constructed the facilitation of economic interests in a positive way that justified a permissive

governance approach, the opposite approach appeared to be favoured within focus groups.

7.6 Reflections and Next Steps

This chapter has presented the qualitative findings that shed light on participants' problematisations of the reuse of administrative law to facilitate a comparison with the problematisations that prompted the GDPR and DEA. The analysis sheds light on how the participants seemed to be more cautious and uncertain of the potential benefits, which led to different problematisations of the governance of the reuse of data in research. The problematisations presented in this chapter were less focused on the opportunities presented by technological advances. Whilst no single problematisation was agreed upon by participants, key themes regarding low levels of trust in the actors involved and uncertainty as to the ability of research to respond to societal challenges were identified. In exploring and discussing their views, participants drew upon external events that undermined their trust and left them uncertain about how much they would benefit from this data processing practice.

The next chapter will build upon these findings to present the qualitative findings that explore the effects generated from the problematisation identified in this chapter. It will explore how the problematisations identified in this chapter and the factors that shaped them prompted certain responses regarding how to govern data reuse in research. To do so, the next chapter will draw upon Bacchi's ideas on discursive effects and subjectification and the notion of legitimacy as a quality underpinned by a system of socially constructed norms, values, beliefs, and

definitions. These findings will complement the findings presented in this chapter to shed further light on the divergences between the GDPR, the DEA, and participants and the potential underlying causes of these divergences.

8 QUALITATIVE FINDINGS II: THE EFFECTS GENERATED FROM THE FOCUS GROUP DISCUSSION

8.1 INTRODUCTION

This chapter builds upon the previous chapter's findings, which explored the potential factors that shaped participants' problematisations of the governance of the reuse of administrative data in research. The focus here is upon the 'effects' generated from the prominent discourses. This sheds light on the 'solutions' participants deemed necessary to render reusing data in research an appropriate activity. Bacchi's[1] framework influences this approach by highlighting how certain problematisations generate certain effects arising from the 'solution' to a particular problematisation. The discursive effects refer to what is rendered thinkable and unthinkable, appropriate and inappropriate through the dominant discourses. The subjectification effects refer to how people are thought about in response to the dominant discourses that shaped the problematisation and the resulting solution. The legislative practices analysed in Part I could identify dominant discourses that engendered certain effects. In contrast, the qualitative data suggested that participants' diverse views have generated, at most, prominent discourses with no consensus, leading to a range of potentially clashing effects.

The chapter's key findings suggest that the focus groups' prominent discursive effects diverge from those identified in Part I. The discursive effects of

[1] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

the GDPR and DEA promoted a permissive approach to reusing data in research by constructing research as an activity in the public interest and the technological developments associated with data processing as opportunities not to be missed. The discursive effects justified a permissive governance approach as appropriate. In contrast, the discursive effects generated by the focus groups are less optimistic and more cautious, questioning the appropriateness of research and the opportunities presented by technological developments. Considering these effects, many participants desired more individual control and a greater role for anonymity than present in the GDPR and DEA to render the reuse of data appropriate. The subjectification effects generated by the GDPR and DEA construct data subjects as empowered and supportive of research do not align with the qualitative findings. The prominent subjectification effects arising from the focus groups constructed participants as resigned, confused, and concerned about the potential consequences arising from research that reuses administrative data.

The qualitative findings further emphasised the diversity of views held by participants, which is a key finding of this thesis. The range of values, beliefs, norms, and definitions identified in the previous chapter created a range of discursive and subjectification effects. This chapter focuses upon the effects of these diverse views, questioning how the law can demonstrate legitimacy in this context where it is impossible to identify a single system of socially constructed views, norms, and definitions. The empirical findings suggest a range of diverse, sometimes conflicting, and sometimes nascent norms, values, and beliefs underpin what participants construct as appropriate in this context. Part III further

explores the challenges for the law arising from this diversity and the potential responses.

8.2 THE DISCURSIVE EFFECTS GENERATED WITHIN THE FOCUS GROUP DISCUSSIONS

Drawing upon Bacchi's[2] work, the discursive effects in this chapter refer to what participants constructed as appropriate and desirable in the context of reusing administrative data in research. These findings help identify potential divergences between what participants viewed as appropriate and what is legally permissible. This section explores divergences concerning the level of individual control, the role of consent, the construction of research as an activity in the public interest, and perceptions of the law's ability to reassure participants regarding the appropriateness of an activity. The findings suggest that participants found a less permissive approach towards the reuse of data to be appropriate than reflected in the current approaches of the GDPR and the DEA.

8.2.1 That Individual Control Should Play a Greater Role

One key divergence identified by comparing the discursive effects identified in the empirical data analysis and Part I's analysis of the GDPR and DEA is the role of individual control. Neither the GDPR nor DEA requires the active participation of citizens to render reuses of data in research lawful. Part I attributed this approach to the discursive effects generated during the legislative processes. These constructed research as a beneficial activity, providing unmissable opportunities that justified a permissive governance approach without the need for individual

[2] Ibid.

involvement. In contrast, in the focus groups, the prominent discourses questioned the benefits of such research and doubted the benefits of advances in technology. This led to a less permissive approach than reflected in the law being deemed appropriate by many participants, who felt there needed to be some form of involvement by individuals.

To some, the need for an individual's involvement was expressed through the notion of consent. Some participants constructed consent as necessary to make the reuse of data in research an appropriate activity. In contrast to the findings in Part I, many participants questioned the benefits of research and instead identified it as a risky activity. As a result, many participants felt it was appropriate for individuals to have some level of individual control to ensure their data would not expose them to risks. Identifying an activity as 'research' was insufficient to make participants feel it was appropriate for their administrative data to be reused for this purpose. Instead, many were more nuanced in their assessment.

In response to a provocation that permitted the reusing administrative data in research without consent, some participants felt that consent was required to make any data reuse appropriate. P3FG2 stated, 'I think that should be made illegal and against the law - rather than us giving information, you have to ask for our permission.' P5FG4 echoed this sentiment, feeling that 'it is important for them to ask our permission before they start this research.' The discursive effect of consent as a necessary aspect of rendering reuses of data appropriate to participants diverges from the legislative approaches that permit data to be reused without consent. For other participants, the discursive effect generated was that

consent was only required to render the reuse of identifiable data in research appropriate. P4FG2's comments reflect this through

> *the idea that I'm identifiable for something that I haven't necessarily consented to, even though there may be instances that's happening right now, and I don't know about it but it's the thought of it that does make me very much feel uncomfortable.*

This chimes with P4FG4's view that 'regardless of what kind of data is being used, actually without having some kind of consent with actual people, they wouldn't want to be as easily identifiable...' These findings align more closely with the approach of the DEA than the GDPR, as the DEA only permits the reuse of de-identified data. This supports the argument that participants are more selective in their support regarding reuses of administrative data in research. This led to a less permissive approach than present in the GDPR being deemed appropriate. Whilst aligning closer with the DEA, the findings suggest the effect of only research using de-identified data warranting a permissive governance approach, is underpinned by different beliefs and values. The approach of the DEA was attributed to its construction of quantitative research as the type of research that justifies the permissive governance approach created by Chapter 5, Part 5 of the DEA. Quantitative research is less likely to require identifiable data to fulfil its purpose, and so this type of research is more likely to be able to make use of the permissive legal gateway for sharing administrative data that the DEA created. Quantitative research, however, was not so positively received by participants. However, participants' views may be due to the influence of the type of data being

linked, rather than the types of research. The empirical work of Aitken and others[3] suggested that the type of data linked was the most influential factor in shaping participants' preferences for the use of data in research.

Not all participants agreed that consent was necessary to make reuses of data in research an appropriate activity. Some participants recognised individual consent could cause practical issues, making consent inappropriate in this context. P4FG4 felt that 'ideally I would want to be able to give consent and permission to use my data but, realistically I don't think that would be feasible, like pragmatic for researchers to do at such scales.' However, the effect of consent not being deemed appropriate in this context did necessarily result from the view that research was an activity in the public interest. Participants' responses differed, suggesting a lack of consensus as to whether research was an activity that was sufficiently beneficial as to be worthy of access to administrative data without consent. These differing views on what is appropriate are reflected in the exchange between P3 and P6 in FG2.

> *I think it'd be a lot of admin, a lot of hassle, and maybe I'd get fed up of looking at all where but also I think that I have a right to know so, yeah, I would say although maybe it's not practical, it's the right thing to do [P3FG2]*
>
> *I think from a practicality point of view, I'd just be more comfortable with it if there were specific uses that were outlined to me that the government couldn't do with it rather than maybe them needing to contact me every time they did process or access my data. [P6FG2].*

[3] Mhairi Aitken and others, 'Public Preferences Regarding Data Linkage for Health Research: A Discrete Choice Experiment' (2018) 3 International Journal of Population Data Science <https://ijpds.org/article/view/429>.

For others, the appropriateness of reusing data without consent was not premised upon the worthiness of research as a beneficial activity. Instead, some expressed a desire to avoid an overly burdensome chore that P2FG3 compared to be continually cold-called. The findings, therefore, provide some support for previous research. Aitken and others' study[4] suggested that citizens recognise that the collection of consent in this context may be problematic and the advantages of individual control may be outweighed by the disadvantages of preventing research taking place. Here, whilst several participants did recognise the practical issues, different factors helped inform the balance of what was appropriate beyond the disadvantages of research not taking place, such as the burden placed on individuals.

During the focus groups, some participants developed their views towards the role of individual control in determining the appropriateness of reusing data in research during the discussion. For example, the comments made by P3FG2 in some instances suggested that, to them, consent was a pre-requisite. During the focus group, their comments suggested that what was important to them was being aware of what could be done with their data, and consent facilitated this. Towards the end of the discussion, they stated that 'I don't know how I would like more control but certainly it would be a start to know how my data was used.' The comments demonstrate how prominent discourses in the focus groups served to refine what participants deemed appropriate in this context.

[4] 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 <https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-016-0153-x#Fn72>.

P3FG2's shift away from requiring individual consent resonates with previous research. Aitken and others [5] found that participants within their research displayed similar shifts in their views. They premised these changes upon the interaction of several factors, including participants' views towards the value of research and their trust in institutions involved in processing their data. This is echoed by the Wellcome Trust's research that suggested that as participants become more aware of what happened with their data, they become more 'open and trusting in the decision-making process around data collection and sharing.'[6]

Such findings reinforce the importance of societal views in this context. They suggest that alternatives to the traditional use of consent may be sufficient to render the reuse of data in research appropriate, where they foster an awareness of how data may be used or enable another opportunity for some form of individual input. Drawing upon Poort's[7] framework for analysing the quality of a law, she argues that societal problems, to which the law should respond, are formed within a given society, shaped by a society's views and values. Where the law ignores or diverges from the societal context, it risks failing to respond to the problem to which it ought to be the legislative solution. Several participants' comments suggested that the permissive approach to governing reuses of

[5] Ibid.

[6] Wellcome Trust, 'Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data' (2013), 13 <https://wellcomelibrary.org/item/b20997358#?c=0&m=0&s=0&cv=0>.

[7] Lonneke Poort, 'Consensus & Controversies: An Interactive Legislative Approach to Animal Biotechnology in Denmark, Switzerland and the Netherlands' (PhD thesis, Tilburg University 2011).

administrative data in research was inappropriate. This inappropriateness was due to the participants' understandings of the risks of allowing such research to occur. Doing so suggests that the law's current approach may fail to respond to the challenges that participants believe exist due to the reuse of identifiable data. Many participants felt such research was only appropriate where anonymous data were used and were unhappy at the thought of researchers being able to reuse their identifiable data, even where participants acknowledged the benefits of such research. This supports Poort's arguments[8] regarding the law's fit with social context being a critical aspect in shaping citizens' views towards the law's legitimacy.

At the same time, however, the findings highlight the nascent nature of the beliefs, views, norms, and definitions that guided participants' views towards what was appropriate. Within the focus groups, many participants altered and developed their views during the discussions in response to others' points or because of their own beliefs and values. Some participants completely changed their minds towards aspects of reusing administrative data in research. Other participants sometimes struggled to reach a definitive conclusion about how they felt about different aspects of the provocations. This raises questions about the law's ability to demonstrate legitimacy as the data undermines the claim that it is possible to determine what is appropriate by reference to a socially constructed system of norms, views, beliefs, and definitions in this context.

[8] Ibid.

A small minority found it appropriate for their data to be reused in research from the start of the focus group. However, these views did not become prominent during the focus groups and thus failed to influence others' views towards the appropriateness of reusing administrative data in research. Therefore, the impact of these views on what participants found to be appropriate remained minimal. P8FG2 justified their view on the basis that consent was no longer required where it had been collected by the government as, at that point, it had ceased to be 'their data.' Other participants disagreed with this view. P7FG2 felt they still held a 'vested interest' in their data and P4FG2 argued the GDPR served to bring 'back power to individuals on their, on what is their information and their data and giving back control.' P3FG4 felt that consent was unnecessary as they had 'nothing to hide' and, as a business owner, 'that information's available for people to look at… cos when we get new contracts, new business, the first thing people look at is Companies House.' When probed further on their views, P3FG4 acknowledged that they were very open about their data but 'it's different really for each person.'

The previous chapter's findings suggested that participants' problematisations of the governance of reusing data in research diverged from the legislative problematisations identified in Part I. This section presents findings that indicate these diverging problematisations enabled different discourses to become prominent. These prominent discourses generated diverging constructions of what is appropriate in this context compared to the discourses identified in Part I. For example, the discursive effects from the legislative processes constructed research as an activity in the public interest that warranted a permissive governance approach. In contrast, the prominent discourses within the focus

groups were more focused on the risks posed by research and technological developments. This prompted a less permissive governance approach and a greater role for the individual within the focus groups. This is not the only example of participants' constructing a more restrictive approach as appropriate. The next section will explore the findings surrounding the role of anonymisation and its role in making the reuse of administrative data in research an appropriate activity to participants.

8.2.2 The Need for Anonymity

The prominent discourses within the focus groups suggested participants felt a greater use of anonymisation than required by the GDPR and DEA was necessary to render reuses of administrative data in research appropriate. The GDPR delineates between different types of data, including personal data, [9] pseudonymous data,[10] and special categories of data.[11] It adopts a permissive approach to reusing these data types in research than other data processing activities. Reusing data for research purposes benefits from a presumption of

[9] Article 4(1) GDPR defines personal data as *any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, 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.*

[10] Article 4(5) GDPR defines pseudonymous data *as personal data that 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.*

[11] Article 9(1) GDPR lists personal data revealing *racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation as falling within the scope of special categories of data.*

compatibility, which does not require researchers to find a separate ground to justify their processing.[12] Furthermore, though special categories of data can only be processed where there is a legitimising ground, research is listed as an example.[13] The DEA appears to be less permissive, permitting the reuse of de-identified data only.[14] Nevertheless, the Explanatory Notes to the GDPR acknowledge that both 'non-personal and personal information can be disclosed pursuant to section 64(1)', but it must be processed so that it is 'not *reasonably likely* that the identity of a person will be deduced from the information (whether by itself or taken together with *any* other information).'[15][16]

These legislative approaches shed light on the discursive effect of research being defined as an activity in the public interest. The particularly permissive approach within the GDPR towards the governance of reusing administrative data in research suggests research is constructed as an activity particularly worthy of such an approach. This enables researchers to reuse personal and sensitive data. The DEA's approach suggests that research as an activity in the public interest is defined differently. However, the scope of the data that can be shared due to the DEA's legal gateway is not confined to anonymous data. The difference in the

[12] See Article 5(1)(b) GDPR that states *further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes.*

[13] See Article 9(2) GDPR for the ways in which a data processor can disapply the prohibition on processing special categories of data.

[14] See s64(3) DEA 2017 that requires data to be processed before it is disclosed so that (*a) the person's identity is not specified in the information, and (b) it is not reasonably likely that the person's identity will be deduced from the information (whether by itself or taken together with other information).*

[15] Explanatory Notes to the Digital Economy Act 2017.

[16] See Article 4(5) GDPR.

definition of what is appropriate to share with researchers may be due to the high value the DEA attributed to quantitative research, which is less likely to need identifiable data.

The previous chapter identified the prominence of certain focus group discourses that did not construct research as an activity in the public interest to the same extent as the dominant discourses identified in Part I. In contrast to the discourses generated through the legislative processes, participants were more focused on the potential risks of reusing data. Rather than activity in the public interest, several participants constructed the potential for identifiable data to be reused in research as overly intrusive, even where it could be beneficial. The problematisations within focus groups thus led to different approaches towards potential 'solutions' that would render this use of data appropriate than identified in Part I. In determining what would be appropriate, many participants prioritised their individual privacy interests and felt anonymisation had a key role to play. This is less permissive than the approach of the GDPR, as reflected in P5FG3's comment that

> *I'm sure it's for the right reasons to help but errrm in terms of the very personal identification, I'm not completely happy because it is too, too personal and specific really and I just think, just think, it's kinda, errrrm I, I don't know, hinders isn't the right word but a certain amount of your personal privacy and even though there's not saying there's anything to hide, it's just that the amount of personal privacy you would really just like to retain*

This approach aligns closer to individuals' views collected during the public consultations analysed in Part I. This supports the argument in Part I that subsequent practices by more powerful actors dominated these views and that

individuals' values, beliefs, norms, and definitions, which shaped their assumptions about what would be appropriate in this context, had a negligible impact on the final legislative texts. This more pessimistic construction of research, which in turn only justifies access to anonymous data, was a point of agreement in FG1.

> *In terms of the data, yeah, I still wouldn't be comfortable with sharing my data if I could be identified as well [P4FG1]*
>
> *I would just agree, if I was identifiable but apart from that I would be very happy to share [P6FG1]*

The qualitative findings suggest that the discursive effect generated from the focus groups aligns closer with the approach of the DEA, which only permits the reuse of de-identified data. The findings that suggest reusing data pertaining to individuals' names were regarded as particularly inappropriate support this argument. P4FG2 stated that

> *I don't necessarily have a problem with that in general, whether, whose, where that information goes when it's just aggregated information, it's not identifying, like there's not names, there's not that information where people can collate it and go 'ohh there's X Y and Z...' and identify people.*

P1FG1's criteria for determining what was appropriate in this context echoed this view. P1FG1 felt that 'if they're just discussing demographics of certain groups, that's okay but I wouldn't want my name or personal information revealed.' Despite the focus group findings aligning closer with the approach of the DEA than the GDPR, there are two caveats. Firstly, it remains unclear how far participants' definitions of not being identifiable align with the definition of de-identified data within the DEA. The closer this definition is to genuinely anonymous data, the further participants' views on what is appropriate diverge

from the DEA. Secondly, the findings suggest the underpinning values, norms, beliefs, and definitions that led participants to deem the reuse of non-identifiable information as appropriate diverges from what underpins the DEA's approach. In contrast to the dominant discourses during the creation of the DEA, which suggested qualitative research was particularly worthy of a permissive governance approach, the previous chapter demonstrated that several participants disagreed with this. The 'solution' of only permitting the reuse of de-identified data was similar, but it was premised upon participants' views on the risks prompted by such research and low trust in the actors involved, rather than a preference for research that requires de-identified data. This finding demonstrated the usefulness of Bacchi's framework in further unpacking the different motivations and reasons that may prompt similar effects.

Within the focus groups, the definitions of research suggested by participants' comments diverged from the definition suggested by the permissive governance approaches of the GDPR and DEA. Participants in the focus groups preferred a less permissive governance approach, which permitted only the reuse of anonymous data, to define data reuse in research as appropriate. Rather than constructing research as a beneficial activity in the public interest whose innovative nature could help resolve societal challenges, participants' definitions were more ambivalent. Comments made by several participants reflect this ambivalence, who regarded the benefits and opportunities as insufficient to justify the reuse of identifiable data. P2FG4 stated, 'I think the objective sounds good, but I think…I would prefer or, you know, want to know, that they would be able to anonymise the data.' P4FG1 felt that

obviously it's a really good research and like, it's, it.... its' for like the betterment of the community but I'm just a bit put off about the amount of data that you would be giving out about me, especially when you can link it back to myself.

The participants from a minority ethnic background appeared to be particularly concerned about researchers having access to identifiable data and data disclosing their ethnic origin and religion. The prominent discourses within the focus groups focused on these risks, generating the effect of anonymity being part of the response to enable only appropriate reuses of data in research. This highlights how participants' own backgrounds may shape their views towards the appropriateness of this data processing practice. Doing so illustrates the difficulty of constructing society as a monolith with just one view. P3FG2's comments suggest their desire for control stemmed from a need to know what would happen to their data and where it would go because:

I'm very easily profiled, I wear a hijab and I'm you know, hah, I feel like specifically for me, I don't want anyone profiling me, like, just as easy as you can do just because of the colour of my skin, or because of how I dress or and, and my life decisions.

Within a different focus group, P4FG3 echoed these sentiments.

my main concern is because I am from an ethnic origin which is not British, that's my problem because…there can be biases… anything, anywhere, you fill any form they ask you, you do any survey they ask you what's your ethnic origin and I always think to myself, why do they want to know my ethnic origin, you know, what's the reason? Because I'm not British, I think about this and that's the only thing that bothers me.

Participants, therefore, did not construct quantitative research as an activity in the public interest but problematised such research as an activity that risks discrimination against them. The beliefs generated the effect only the reuse

of anonymous data in research being constructed as appropriate. The desire for further anonymity as part of the 'solution' to the governance of reusing administrative data in research aligns with previous research that explored individuals' views towards the reuse of data in research.[17]

The prominent discourses within focus groups diverged from those within the legislative processes to highlight and emphasise the risks and challenges posed by this use of data. This more negative problematisation thus justified a less permissive solution than that of the GDPR, which permits the reuse of identifiable data. While the DEA aligned closer with some participants' views by only allowing the reuse of de-identified data, this approach did not respond to the concerns of all participants. Some remained concerned about the potential for the reuse of data in research to foster discrimination. The previous chapter's findings and this section suggest that the views and assumptions constructed 'research' in a less favourable light than found in Part I. The following section will detail the effects generated towards research as an activity in the public interest.

8.2.3 Questioning Research as an Activity in the Public Interest

The analysis in Part I suggested the practices during the legislative processes constructed research as in the public interest. This understanding of research thus justified a permissive governance approach towards the reuse of data in research within the GDPR and DEA. The empirical findings suggest that participants did

[17] See, for example, Cameron and others, 'Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes' (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/> and Gillian Robinson and others, 'Public Attitudes to Data Sharing in Northern Ireland: Findings from the 2015 Northern Ireland Life and Times Survey' (2018) <https://pure.ulster.ac.uk/en/publications/public-attitudes-to-data-sharing-in-northern-ireland-findings-fro>.

not construct research so positively and thus did not find such a permissive governance approach appropriate. Instead, the prominent discourses were ambivalent towards research generating an effect of research as not always an activity in the public interest.

The focus groups did not provide a consensus as to the definition of the 'public interest.' However, the prominent discourses questioned the benefits arising from research reusing administrative data in several ways. For example, the prominent discourses suggested that research creating financial benefits or supporting economic interests were deemed contrary to the public interest. As a result, many participants felt the inclusion of commercial entities rendered reuses of administrative data inappropriate. Furthermore, several participants then questioned the appropriateness of academic entities in this context. In contrast to the 2015 Conservative Party Manifesto, which constructed the building of relationships between industry and academia as a positive development, several participants felt the increasing commercialisation of academia rendered academics' involvement as potentially inappropriate and not in the public interest.

The prominent discourses did not just question the public interest facilitated by research that supported economic interests. They also questioned the ability of research more broadly to be in the public interest. This more generalised concern arose through discussions that proposed research could enable the 'targeting' of certain groups and allow actors to make 'assumptions based off of linking two things together' [P1FG4.] For some participants, this discourse had the effect of them doubting that research could be in the public interest. P4FG4 stated this was

> *why I brought up the thing about bias because it just comes up over and over again and it's just one of those things that you try to reduce as much as you can, but you can't ever completely eliminate it.*

A discussion of the provocation concerning the reuse of administrative data to respond to employment issues demonstrates the discursive effects of what is perceived as appropriate. P4FG1 felt

> *there is a possibility where, like, those who are using those types of schemes, won't benefit as much compared to other schemes that the government uses in the future so there is that possibility of ermmmmm the research not really benefitting other people as well when it should have been a win-win situation for everyone.*

The prominent discourses within the focus groups suggest a construction of research that is less inherently positive than that proposed by the dominant discourses that shaped the GDPR and the DEA. Participants were less focused on the positives of research and innovation and were more uncertain about the outcomes of research reusing administrative data.

Furthermore, the prominence of the questioning of research as an activity in the public interest prompted participants to find research outcomes relevant to their assessment of the appropriateness of research. The discursive effect that the outcomes of research ought to help inform the appropriateness of the activity diverges from the legislative approaches. The problematisations identified in Part I elided the activity of research with the positive outcomes of research and suggested the latter would inevitably arise where research took place. This elision served to justify a permissive governance approach as appropriate. The focus group discussions did not provide support this elision. Instead, participants' definition of research that justified the reuse of administrative data emphasised

the importance of outcomes to their definition of research as an activity in the public interest. There were, however, differences between what outcomes participants felt would make the reuse of data in research an appropriate data processing activity.

Several participants' comments suggested they defined research as an activity in the public interest that made the reuse of administrative data appropriate when it provided tangible benefits. For example, P4FG5 felt they 'would just be really annoyed if I found out someone had gained access to my records and then it didn't actual inform anything.' Furthermore, P4FG2 felt that 'if we're not gunna see local changes or investments into us there needs to be the direct compensation in some way.' There were, however, competing discourses regarding the relevance of research outcomes in defining whether this research was an activity in the public interest. Some participants did not necessarily link the outcome of research with its ability to be potentially an activity in the public interest. For example, P5FG2 acknowledged that 'you can't be assured of the outcome, so you have to be open to it, maybe not necessarily bringing about those outcomes.' For others**,** the increase in knowledge accrued in the academic sphere seemed sufficient, albeit not an ideal outcome. P1FG3 stated

> *I mean, the researcher will probably be, I mean, they'll end up with a PhD which I suppose is, which is a good thing errrm, if its, if its, just a paper exercise ermmm, I don't know, it's a difficult one, it's a difficult one errrrm…*

Participants drew upon their own experiences of participating in research to determine what was in the public interest in this context to support their views. For P3FG2 their experiences in participating in the 2021 National Census made

them question whether reusing data in research was in the public interest. The inability to opt out of the Census, the burden that it placed upon them, and the lack of any noticeable benefits arising from their participation undermined the value they placed on research. Other participants drew upon their personal experiences to construct their definition of research as an activity in the public interest. P2FG3's comments demonstrate the significant impact of a participant's experiences towards shaping their views about the appropriateness of reusing data in research

> *I said, from a personal level and under my current circumstances, I would be happy if the government was to reach out to me and say you know we'll send you on courses or whatever. If you'd have asked me 5 years ago when my life was completely different, I'd, I would have found it a bit intrusive.*

These findings suggest a tension between the quantitative and qualitative findings of this thesis. The prominent discourses within focus groups suggest a more restrictive definition of research as an activity in the public interest than suggested in the qualitative findings. As discussed in Chapter 6, most respondents found research that increased existing knowledge to be an appropriate reuse of administrative data in research. To explore this further, I analysed the survey responses of the focus group participants. Of the 36 individuals that participated in the focus groups, 63.9% felt it was either very, or quite acceptable for administrative data to be reused in research where the outcome was the further knowledge, rather than to provide benefits to either themselves or others in society. Such support did not translate to a prominent discourse within the focus groups. Instead, the discussions suggested that research needed to generate a tangible benefit for participants to define it as in the public interest.

The importance of research outcomes in defining whether it is an activity in the public interest, generated by prominent discourses that questioned its ability to benefit society, diverges from the approach of the GDPR. The GDPR's permissive approach is premised upon a definition of research as an activity in the public interest irrespective of the research outcomes. The DEA's approach may align closer to the effects generated by the prominent discourses within focus groups that suggest research is in the public interest where it provides tangible benefits. Research projects must demonstrate they are in the public interest as one of the conditions for benefitting from the legal gateway created by the DEA. Nevertheless, it is unclear how far the definition of the 'public interest' as a criterion within the DEA aligns with the definition proposed by the prominent discourses within the focus groups. Chapter 4 highlighted the practices that shaped the DEA's legislative process, which constructed the linkage of industry and academia as a positive development. The findings suggested that the definition of the public interest was constructed to best suit the needs of public authorities to respond to what they felt was appropriate and important. This diverges from the prominent discourses that shaped the public interest in focus groups. The inclusion of commercial actors was constructed by many as undermining the public interest that could arise from research reusing administrative data. Furthermore, several participants doubted whether public authorities would act in line with their definition of the public interest and instead felt research would be used to justify pre-determined decisions.

The findings suggest that the definitions arising from the prominent discourses generated a more fine-grained approach towards distinguishing which

research activities ought to be able to reuse administrative data. An interaction between one participant and the facilitator regarding their views towards research as an activity potentially in the public interest demonstrates this approach. P1FG2 found such research to be 'risky but good - I'd probably classify it as….' When asked by the facilitator whether they viewed it as 'worth the risk?' They responded by saying 'errrrrrr, I'll reserve judgement on that on a project-by-project basis.' The doubts as to whether research is an activity in the public interest meant participants constructed greater control and increased knowledge as necessary to render the reuse of data in research appropriate. These findings cast doubt on the legitimacy of the law governing the reuse of data in research as they suggest a broader category of research may benefit from a permissive governance approach than participants would find appropriate. This divergence stems partly from differences in the definitions of research in the public interest, which justifies this permissive governance approach. The next section will explore the impact of such divergences on participants' assumptions about the law's ability to reassure them of the appropriateness of reusing data in research.

8.2.4 Legality Fails to Reassure….

The prominent discourses identified in the previous chapter generated the effect that many participants found the legality of research insufficient to define it as an appropriate reuse of their data. Divergences between what is legally permissible and what participants found to be 'appropriate' question the legitimacy of the legal governance of the reuse of data in research by suggesting the values, beliefs, definitions, and norms that dominated the legislative processes are not dominant within society.

This law's inability to shape participants' definitions of research reusing administrative data as appropriate is reflected in the comments of P4FG2. They stated that 'if it becomes legal, it is was it is, but I wouldn't necessarily be happy with it, it wouldn't really reassure me.' A prominent discourse within the focus groups that helped generate this effect concerned several participants' mistrust in the legislative process. For some, they felt such processes failed to account for the interests of citizens properly.

> *I just don't see how many people would be happy about this and it's just another example of the government just doing what they want regardless of people's opinions. It's not taking account of how people feel about how their data's used and I don't think it would go down well at all. [P8FG3]*

This effect arises from the previous chapter's findings, whereby the prominent discourses constructed the government as an actor insufficiently concerned with society's best interests. This discourse, supported by the participants' discussion of high-profile political scandals, including the Greensill lobbying scandal,[18] generated the effect of law being deemed a tool for enabling the government to help certain individuals rather than broader societal interests. The scandals were not directly relevant to reusing administrative data in research but still generated the effect of participants constructing a restrictive approach towards this activity as appropriate due to this mistrust.

Prominent discourses that expressed frustration over how little input citizens had in decision-making further contributed to the feeling that the

[18] Danica Kirka, 'UK Lobbying Scandal Snares Ex-PM Cameron' *AP NEWS* (April 25, 2021).

lawfulness of reusing data was insufficient for participants to define it as appropriate. P3FG2 stated

> *it doesn't sound like we'd have a choice but to kind of go along with it, it's not something that I'd be ermmm happy to do....ermmm...but....I mean, it doesn't sound like, I don't know what I can say to that , it's not something I'd be happy with but ultimately I know that it's something that I have to do because its legal, it's legally binding so...*

These findings suggest that participants' views of the law being insufficient for participants to assume an activity is lawful appears to be underpinned by their definitions of the law. The definition of the law suggested by the GDPR and DEA's approaches constructs the law as a tool to facilitate beneficial activities in the public interest, such as research. In contrast, the definition of law suggested by participants' comments constructs law more as an authoritarian tool that rode roughshod over their beliefs.

Other views shared during the focus groups have a similar effect of suggesting the legality of an activity was insufficient for participants to define it as appropriate. Alongside the view that the government was more concerned with promoting the interests of a certain few rather than the best interests of society were prominent discourses that criticised the lack of representativeness within government. Participants found the lack of diversity within government as a reason why the government's approach to what is legally permissible diverged from participants' views on what would be appropriate in this context. P3FG2 stated that

> *I feel they're slightly detached from the actual issues in local communities umm so...they make up the government and they come*

> *from areas, I feel like they can be slightly detached so I'm not sure it's a true representative of me and my current issues and my circumstances who are sat in government.*

P3FG2 subsequently lamented

> *there's not many Asians, Black, ethnic minorities that are representative of the different cultures in government […] I haven't seen a representative version of me, you know, I don't think that exists…*

P4FG2 strongly supported this sentiment, feeling that such decisions needed to be made in a 'setting with a diverse range of people which make up like the fabric of society.' The effect of the prominent discourses, which meant the legality of reusing data in research failed to reassure participants of the appropriateness of this activity, resonate with the findings of Carter and others.[19] Their study focussed on the public concern expressed over the care.data data sharing initiative.[20] The divergences between what those in power believed the public would support, due to the potential benefits and perceived low risks identified in relation to the public response to care.data, are reflected in these findings. The following section will explore how participants responded to the disconnect between what is legally permissible and what they found to be appropriate.

8.2.5 …Therefore, Something Needs to Change

The prominent discourses not only suggested that the legality of an activity was insufficient for it to be defined as appropriate for participants, but they also shed light on what participants believed to be the appropriate governance approach.

[19] Pam Carter, Graeme Laurie and Mary Dixon-Woods, 'The Social Licence for Research: Why Care.data Ran into Trouble' (2015) 41 Journal of Medical Ethics 404.

[20] Ibid.

This section explores how participants 'solved' their problematisation of reusing data in research as an activity defined as riskier and less beneficial than the definitions used to determine the legislative responses analysed in Part I.

Several participants suggested there was a need for citizens to play a more prominent role in the pre-legislative stages to ensure only appropriate reuses of data were made lawful.

> *I would be unhappy with that because if it's the law, it's not something you can stop so I would be unhappy and maybe errrm if there if it was the law, if there was like people's opinions that could be obtained before, that would be a good idea rather than straight away making it into a law... [P6FG3]*

This suggests P6FG3 felt the research could be rendered lawful and appropriate through the inclusion of citizens into the pre-legislative space as subsequent involvement would be too late to enable the necessary changes. Part I demonstrated that such practices already occur, but the findings and empirical research suggest that the current approach fails to involve citizens sufficiently. The findings suggest this risks the passing of laws that permit activities that citizens define as inappropriate by failing to adequately incorporate a societally constructed system of norms, beliefs, values, and definitions, which may undermine the legitimacy of such laws.

Other participants considered the changes that citizens could make happen in the post-legislative space. The discourses here emphasised the importance of societal acceptance to the proper functioning of the law by reference to the severe impact of previous public backlashes. P5FG3 drew upon their experiences and knowledge to argue that

people can make a difference if they're not happy. I don't think it's a, a situation where we just have to put up with it. If you think of like examples like... I don't know, the poll tax riots, people can make a difference. I know that's a bit extreme example, but it was a situation where the iron fist came down people and people were like 'no, I'm sorry, I'm not gunna put up with that' and it changed the law.

P1FG3's comments echoed this view, feeling that this was

how change is made in society, it always begins with ermmm usually, middle class university educated people saying 'well this situation is very wrong, and we need to do something about it' gathering around them the people that really affects and ask government nicely to do something about it and of course, generally 99.9 times out of 100 the government ignore them. Then the asking gets a little more fervent, then we move onto direct action and then the government start listening. It's when it impacts the government, rather than situation purely just impacting those less fortunate in our society, then that's the only way we can make some sort of lasting change but I, no I would be out there on the street, I'm afraid.

Such comments chime with the outcomes of previous data-sharing initiatives, including care.data,[21] which sought to open-up government-held data to other interested actors, including researchers. The strength of the public backlash led to its eventual cancellation, and attempts to establish a similar initiative in 2021 prompted a similar uproar and a hasty shelving.[22]

The prominent focus group discourses suggested that what may be lawful in this context may not necessarily be appropriate. This led several participants to find that greater transparency could be a potential 'solution' to the potential disconnect between what is lawfully permissible and what was deemed

[21] Sigrid Sterckx and others, "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' (2015) 19 Medicine, Health Care and Philosophy 177.

[22] Sara Mageit, 'UK Government Halts GP Data Sharing Scheme after 'Mistakes' Were Made' *Healthcare IT News* (July 22, 2021).

appropriate by participants. Here, the inappropriateness of what is lawfully permissible in this context was attributed not to individuals' lacking control but rather lacking knowledge. The comments suggested some participants found such reuses of data were inappropriate as they violated the norm that people ought to know how their data may be used.

> *I was about to say that rather than more control I was thinking more transparency cos that a more realistic and easier thing that can be done…I just feel it would make people feel better cos it's also easier to hold people accountable if something is used that doesn't align with you in that way because they have been transparent about it beforehand [P4FG4]*

This 'solution' aligns with the prominent discourses analysed in the previous chapter, particularly those defining individuals as the owners of their data and constructing the actors in this context as untrustworthy. To P4FG4, transparency was a 'solution' to what they found to be inappropriate in this context due to its perceived ability to foster better mechanisms of accountability.

However, for other participants, transparency as a 'solution' to the problems identified in the reuse of data in research did not stem from its ability to ensure accountability. Instead, the prominent discourses that focussed on the untrustworthiness of certain actors and the risks of research led some participants to find 'transparency' to be a solution for a different reason. Rather than to enable mechanisms of accountability, which would help make this use of data appropriate, some constructed transparency as a solution to their problematisation of the governance of reusing data in research through facilitating greater individual control over data. P1FG4's response to the above view expressed by P4FG4 reflects this view.

> *I think when, if you're not aware, you can't control it so when you don't even know what's happening, you're not even able to make those choices so just having straight-forward transparency would enable you to have more…more...choice and more...space for you to impart your own decisions on the data.*

These data suggest that participants constructed greater transparency as a 'solution' to their problematisation of the governance of reusing data in research. However, the data suggest that two competing discourses generated this effect. One prominent discourse constructed transparency as a solution because it could foster the levels of accountability required to render this reuse of data appropriate. However, others supported transparency as a solution by reference to its ability to facilitate individual control. This supports the claim that it is only possible to talk of prominent, rather than dominant, discourses arising from the focus groups, which generates a range of discursive effects as to what participants constructed as appropriate. These findings suggest that greater transparency would not necessarily entail individuals constructing the reuse of data in research as appropriate where it did not respond to the underlying problematisation. To P4FG4 the focus is on holding certain actors accountable, which is deemed necessary due to the low trust participants have in such actors, whereas P1FG4 defined transparency as a mechanism that promotes individuals' autonomy. The discursive effects were not the only effects identified within the analysis of the focus group data. The next section will explore the subjectification effects generated within the focus groups to consider the potential divergences to the subjectification effects identified in Part I of the thesis.

8.3 THE SUBJECTIFICATION EFFECTS GENERATED WITHIN THE FOCUS GROUP DISCUSSIONS

This section presents the findings that shed light on the subjectification effects generated by the prominent discourses of the focus groups. The aim is help identify divergences between the effects arising from the focus groups and the subjectification effects of the GDPR and the DEA. This draws upon Bacchi's arguments that certain problematisation engender a particular construction of the actors impacted by policies and laws. These constructions, or subjectification effects, contribute to the understanding of what is the most 'appropriate' solution.

8.3.1 The Citizen as Overwhelmed and Resigned

The subjectification effects generated by the provisions of the GDPR and DEA constructed citizens as supportive of research, justifying their permissive approach towards the governance of the reuse of data for research purposes. The prominent discourses within the focus groups fail to generate the same subjectification effects. Instead, they construct citizens as begrudgingly tolerant at best. P4FG1 comments suggested they felt a certain duty in this context that they were not happy to bear but felt resigned to, nonetheless.

> *[I]n regard to government, even though I have quite strong feelings about it, I'm like, you know, distrustful, not feeling very keen or comfortable with sharing a lot of my data but it is...I can see from the point of view where it necessary for them to have such data.*

Other participants agreed with this view through non-verbal means. P1FG1 echoed this resigned support, acknowledging the potential benefits whilst simultaneously expressing concerns about the management of their data and

mistrust in the government's potential attempts at transparency. P1FG1 did not feel that it would be presented in a way that would be 'visible' to citizens.

In contrast to the subjectification effects generated by the GDPR, which constructed citizens as empowered and active consumers, the prominent discourses in the focus groups constructed citizens as resigned and helpless. In doing so, it suggests any sense of 'support' as to the appropriateness of this activity is underpinned by a sense of begrudging support arising from a lack of alternatives or means of expressing meaningful disagreement.

> *[t]here's a certain amount that's just mandatory by default that's provided to the government to facilitate services they provide back, there's not much you can do about that because its government errrrm well yeah, controlling. [P6FG1]*

This subjectification effect is echoed by P6FG2 who felt they'd 'probably feel quite resigned to the fact that I don't actually have any, like, any, practically I don't have any control on any of this....' Prominent discourses that lamented the amount of data on individuals that is already 'out there' further contributed to the subjectification effect of overwhelmed and resigned citizens. Despite P2FG3 feeling 'uncomfortable' with the reuse of data in research, they were resigned to its existence as 'ultimately if you've got enough insurance, if you've got a bank account, if you've got a postcode, you know, they know this stuff anyway.'

For others, these subjectification effects arose in response to the prominent discourses that constructed a certain relationship between the state and its citizens. In FG2, the participants discussed the appropriateness of reusing data in research by reference to citizens' owning a duty through their citizenship. P7FG2 disagreed with P3FG2's construction of a high level of control and knowledge as

required to render reuses of data in research appropriate. Instead, P7FG2 felt 'consent's kinda given when youuuu say you're going to be a citizen of a place....' The prominent discourses overshadowed the views and norms present in this contribution to the focus groups. Therefore, the prominent discourses in the focus groups remained focussed on the need for individual control to render reuses of data in research an appropriate activity. In response, P7FG2 acknowledged the unique level of power that the government held over citizens' lives due to the wide range of services they alone provided. This uneven yet necessary relationship led P7FG2 to acknowledge the subjectification effect of the resigned citizen, rather than a duty-bound one.

The subjectification effects generated within the focus groups were not universal as not all participants were so resigned. Some participants felt the reforms in data protection law had granted individuals greater power and control over their data and, to P4FG2, there was 'value' in that. Nevertheless, such discourses were a minority and the more prominent discourses within the focus groups remained focussed upon constructing the citizen as resigned and helpless. The comments made by P1FG4 suggest the low levels of trust in the actors involved in the reuse of administrative data contributed to this subjectification effect. To P1FG4, the current approach was suspected to be a deliberate attempt to undermine the ability of individuals to be cognisant as to the uses of their data.

> *I wonder if it's made intentionally complicated, I'm not sure but I think a lot of us are ignorant and that means we don't know what's happening and I question whether that's done intentionally...*

Furthermore, there was a divergence in the subjectification effects generated by the GDPR, where the prominent discourses defined greater control

as a positive development for empowering consumers, and the subjectification effects generated by the focus group discourses. Aligning with the criticisms of the GDPR discussed in Chapter 3, the qualitative findings suggest the construction of an empowered citizen felt by participants to be a façade. For P4FG4, they felt that

> *people are bombarded with so much information from both ends of the spectrum that they just don't really know what to do with it anymore, what to think or to include and that's actually a huge problem for people because, yes, having information out there is good but also, I feel like there's an extent to it because having information overload is not good for anyone.*

Other participants expressed a feeling of being overwhelmed. Rather than a positive development that empowered individuals, some felt the extra control failed to resolve their problematisation of the governance of reuses of data in research. Their problematisation focussed on the potential issues arising from implementing research outcomes rather than control over their data. Therefore, greater control over their data failed to 'solve' certain problematisations and did not respond to their construction of citizens as unable to ensure their data enabled positive outcomes.

> *even if I did have more control over my data, I wouldn't necessarily have control over how that data's used and how the outcome of that data or how the outcome of that research is then used, whether it's used for good or whether it's used for profit or whether it's used for....whatever, you know I'm never gunna have full control over those things so actually...I probably wouldn't be too fussed about having more control over my data because I'm not gunna have full control over what comes out of that data which I think is the most important part for me. [P4FG5]*

These comments chime with previous research that suggest the purpose for which data are used is important in securing public support.[23] In failing to acknowledge this in the legislative approaches, participants constructed individuals as resigned to whatever would happen, rather than supportive or empowered.

8.3.2 Actors as Not Necessarily Acting in the Public Interest

The prominent discourses identified in the previous chapter suggested that many had low levels of trust in the actors involved in reusing administrative data. These discourses fostered subjectification effects that constructed many of these actors as not necessarily acting in the public interest. This subjectification effect prompted a 'solution' within focus groups that was more restrictive than the approaches of the GDPR and DEA identified in Part I.

The previous chapter's findings suggested that many participants constructed academic researchers to be the most trustworthy actor involved in reusing data in research. This subjectification effect arose in response to prominent discourses rooted in participants' beliefs that researchers adhered to high ethical standards and were more neutral than other actors. The prominent discourses constructed researchers as more likely to align with participants' definitions of the 'public interest,' but there was still some discord within the focus groups, preventing a consensus. Certain discourses that portrayed academics negatively due to their increased commercialisation generated the effect of some participants constructing academic researchers to be inappropriate actors in this

[23] Aitken and others (n 3).

context. These discourses included constructions of researchers as interested in making profits and selling their research to the government. This supports the findings of this thesis and previous research that has suggested the pursuit of profit is a 'red line' for many individuals in determining appropriate reuses of data.[24] This diverges from the subjectification effects present in the GDPR and DEA that, due to their market-framing, did not construct the commercialisation of academic interests as irreconcilable with such researchers acting in the public interest.

The subjectification effects generated from the prominent focus group discourses constructed the government as likely to act inappropriately in this context. The prominent discourses that emphasised participants' mistrust in the government and the high-profile scandals that participants drew upon to support their views contributed to this subjectification effect. For others, their assumptions regarding the government as not acting in the public interest arose from their negative construction of the current relationship between citizens and the State, which rendered the citizen helpless and resigned. This suggests that the law governing the reuse of data in research may lack legitimacy as many of the participants did not appear to presume that the government would act in their best interests based on a socially constructed system of norms, beliefs, values, and definitions.

A further discourse that contributed to the subjectification effect of the government not acting in the public interest concerned the lack of diversity within

[24] Mary Tully and others, 'Commercial Use of Health Data—a Public 'Trial' by Citizens' Jury' (2019) 3 Learning Health Systems.

government and its inability to represent the diverse society it governs. P3FG2 felt that

> *you have MPs, you have your House of Commons, you have your different parties that make up err but, but I haven't seen a representative version of me, you know, I don't think that exists.*

P4FG2 supported this view as they agreed that 'you're never gunna totally please everyone.' Nevertheless, they felt greater inclusivity could solve their problematisation of the inappropriateness of the government's role in reusing data in research by contributing to the 'right process.' In doing so, P4FG2 felt

> *if people can have trust in the, the process, like decision making process and know that all groups are kinda accounted for and kinda have a voice in coming to a conclusion in the public interest, which isn't going to appease everyone, no decision can but just knowing that the right process is in place will be, will be enough for people take.*

This suggests participants constructed the government as an actor who was insufficiently representative of the governed to define the 'public interest' in a way participants would deem to be appropriate. In doing so, the participants began to explore alternative and additional governance mechanisms that could shed light on the norms, values, beliefs, and definitions present within society to determine better what should be deemed appropriate and desirable in this context. Nevertheless, not all participants were supportive of such mechanisms, with P8FG2 querying the ability to 'get the public behind one thing' and raising concerns as to the inefficiencies of such a system. The responses of others, who felt that a consensus wasn't necessary but rather the greater inclusion would facilitate more trust in the process, shed light on how the different underpinning norms could lead to different effects. Where efficiency and effectiveness are prioritised,

such mechanisms are seen as burdensome yet, to others, this inclusivity was at the heart of what processes should be used to determine appropriate reuses of data. These divergences, present within focus groups, raises questions as to how the legitimacy of the law could be secured in such circumstances. The conclusion will reflect upon these findings and consider how they will be taken forward into Part III, which explores the final research question. Considering the differences between the legal approach and participants' views, or where a lack of societal consensus undermines attempts to explore the notion of 'alignment', how can legislation and those using legislation to determine appropriate secondary uses of data respond?

8.4 CONCLUSIONS

This chapter presented the findings that considered the effects generated within the focus group discussions. This chapter highlighted that, in contrast to the analysis in Part I, it was rarely possible to find dominant discourses. Instead, competing discourses with different levels of prominence throughout shaped the discursive and subjectification effects arising from the focus groups.

The prominent focus group discourses generated discursive effects on what was considered appropriate that diverged from those identified in Part I. The differences in the definitions of research as an activity in the public interest, the relevance of actors, and the relevance of the research outcomes served to prompt different 'solutions' to the governance of reusing data in research. The positive discursive effects identified in Part I generated a permissive approach as the appropriate response, contrasting with the prominent focus group discourses. The

increased ambivalence in research as an activity in the public interest and mistrust in actors involved in research reusing administrative data prompted a more restrictive solution than present in the law. This was expressed through greater individual control, more anonymity, and changes to the law being deemed necessary to render what is lawful in this space appropriate to participants.

The subjectification effects generated from the focus groups also diverge from those identified as arising from the GDPR and DEA. Many participants' views did not correspond with the construction of the 'empowered citizen' identified in Chapter 3. Nor did participants' views support the construction of the citizen as one entirely supportive of research, as identified in Part I. While the focus groups' views were diverse, the prominent discourses suggested participants constructed themselves as resigned and uncertain and distrustful of actors uphold the public interest. This suggests that the mechanisms designed to ensure appropriate uses of data may not function properly in this context as they are premised upon a construction of a citizen that is not reflected in the empirical findings.

The diverse views present throughout the empirical findings question the law's ability to determine appropriate reuses of administrative data in research. The definition of legitimacy within this thesis has implied the need for a single system of socially constructed norms, beliefs, values, and definitions that the findings have not supported. Nevertheless, the chapter has also demonstrated that many participants expressed dissatisfaction with the current situation and were desirous of change. The next chapter will consider the challenges associated with the governance of activities that prompt a diverse range of responses. Drawing

upon the qualitative findings regarding participants' views towards reform and existing research on the governance of innovative and divisive activities, Part III explores possible responses and the limits of the law to govern the reuse of administrative data in research.

Part III: Going Forward

9 SO, WHAT CAN WE DO?

9.1 INTRODUCTION

The appetite for governments worldwide to invest in big data analytics to provide further insights into societal challenges shows no sign of abating. Indeed, the impact of Covid-19 appears to have accelerated the digitalisation of the public sector,[1] enabling governments to collect greater amounts of data that researchers may potentially access. Nevertheless, these developments are not without their risks. Uses and secondary uses of government-collected data for purposes that citizens may not have expected have prompted several backlashes. The continued pursuit of the greater collection, linkage, and analysis of administrative data risks befalling a similar fate. These backlashes are potentially a symptom of a more pernicious risk – that these continued clashes suggest the legal governance of secondary uses of data in research may lack legitimacy. This loss of legitimacy stems from citizens no longer assuming the data processing allowed by the law is appropriate due to differences in the norms, values, beliefs, and definitions used within society and the ones underpinning the law.

The thesis has examined this in the specific context of secondary uses of administrative data in social science research. Doing so demonstrated a range of potential divergences between the legal approaches toward reusing administrative

[1] See, for example, Department for Digital, Culture, Media & Sport, 'Government Response to the Consultation on the National Data Strategy' (2021) <https://www.gov.uk/government/consultations/uk-national-data-strategy-nds-consultation/outcome/government-response-to-the-consultation-on-the-national-data-strategy> and Institute for Government and others, 'Digital Government during the Coronavirus Crisis' (2020) <https://www.instituteforgovernment.org.uk/publications/digital-government-coronavirus>.

data in research and participants' views on what is appropriate. The analysis in Part I and II suggested potential underlying causes of these divergences and refuted the concept of homogenous societal expectations in this context. But where can we go from here? How can the law governing secondary uses of administrative data and those reliant on legislation to determine when data can be shared respond to the identified divergences?

This chapter posits that attempts to change the laws governing reuses of administrative data in response to the identified divergences with this thesis would be unwise. Instead, drawing upon the work of Edwards,[2] differences between the legal approach and societal expectations are constructed as a 'feature' of the legal system, rather than a 'bug.' A law's enactment is not the endpoint at which no further citizen involvement is possible. Instead, understanding the law as fostering fluid regulatory spaces where a range of actors can participate in different ways is preferred. This enables the development of terms, such as the public interest, to respect better the views and beliefs of those using and those affected by the law governing reuses of administrative data in research.

The definition of legitimacy,[3] theories of responsive law,[4] and literature exploring the concept of a social licence[5] suggest how the identified divergences

[2] Mark A. Edwards, 'The Alignment of Law and Norms: Of Mirrors, Bulwarks, and Pressure Valves' (2014) 10 FIU Law Review 19.

[3] Mark Suchman, 'Managing Legitimacy: Strategic and Institutional Approaches' (1995) 20 Academy of Management 571.

[4] Philippe Nonet and Philip Selznick, *Law and Society in Transition: Toward Responsive Law* (Transaction Publishers 2017).

[5] Pam Carter, Graeme Laurie, and Mary Dixon-Woods, 'The Social Licence for Research: Why Care.data Ran into Trouble' (2015) 41 Journal of Medical Ethics 404.

might be understood as a 'feature' of the legal system. The concept of a social licence concerns 'how the expectations of society regarding some activities may go beyond compliance with the requirements of formal regulation.'[6] These works help propose a dynamic process of norm development that could feed into a developing understanding of the legal governance of reusing data in research. Therefore, to capitalise upon this 'feature', I advocate for meaningful citizen involvement before, during, and after the legislative process.

The chapter explores existing mechanisms where citizens views have been incorporated in novel ways to determine what appropriate governance entails in different contexts. An in-depth investigation beyond this thesis' scope, but these mechanisms provide insights into what may be possible. This chapter, therefore, provides a fruitful foundation for future research in this area. These mechanisms are not a silver bullet to the divergences identified, partially due to the inherent difficulties of an increasingly diverse society reaching a consensus on what is appropriate. However, a focus on substantive involvement, an acceptance of the pluralistic nature of modern society and its impact on our understanding of the 'public interest,' and a shift away from what is preferred towards what is acceptable to citizens, are argued to be critical aspects to such governance.

9.2 Constructing the Law

Part I explored the legislative processes that resulted in two instruments critical to reusing administrative data in research: the GDPR and the DEA. A key finding was that those involved in the legislative process attempted to shape, control, and

[6] Ibid, 404.

demarcate key concepts and categories relevant to the data processing activity of reusing administrative data in research. The analysis, drawing upon Bacchi's[7] constructivist framework, demonstrated that actors' influences were unequal and, in particular, those proffered by citizens were often overshadowed by subsequent dominant discourses. The legitimacy of these legislative outcomes is thus questionable, as the lawfully permissible activities may not correspond to a socially constructed system of norms, beliefs, values, and definitions that shape societal views towards the appropriateness of this data processing. This is despite the latter's apparent importance during the legislative process.

Empirical findings further support the existence of divergence between the legislative approaches and societal views. The findings suggest that it is impossible to construct society as having homogenous views in the context of reusing administrative data in research. Instead, participants draw on different norms, beliefs, views, and definitions in determining what was appropriate in this context. The idea that the law should act as a 'mirror' to societal views has been criticised[8] Tamanaha argues that in pluralistic societies, it 'may well be that presence of a degree of lack of mirroring is a normal state of relations between law and society.'[9] He further notes that there is the possibility of legal systems representing those who control the political and economic systems and does not consider societal norms.[10] As a result, a mismatch between the legal approach as

[7] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

[8] Brian Z Tamanaha, *A General Jurisprudence of Law and Society* (Oxford University Press 2010).

[9] Ibid, 132.

[10] Ibid, 109.

societal views may be 'the best solution for a complex social problem.'[11] Building upon this, Edwards persuasively argues that whilst 'most of the time, a gap between law and social norms does, in fact, place pressure on the law to change to better reflect social norms,'[12] persistent gaps may be justified as a feature of the system, rather than a bug.

Edwards' writings on bulwarks are of relevance to this thesis. Bulwarks 'buttress law against pressure to change, despite a gap between law and social norms,'[13] which aligns with the divergences identified in this thesis. One category of bulwark against legal change is 'the capture of the mechanisms of change by those highly motivated to prevent it.'[14] This 'bulwark' may be present in the governance of reusing administrative data in research, where Part I identified the dominance of discourses that did not align with the views proffered by individuals during the public consultations. The empirical findings in Part II support this argument as they suggested ongoing divergences and different assumptions as to what was appropriate in this context. Edwards also defines 'pressure valves' as mechanisms that 'relieve pressure on the law to change to reflect norms' where behaviours are 'legal but socially unacceptable.'[15] Whilst he cautions against such pressure valves where they may prompt selective enforcement in a discriminatory

[11] Ibid, 132.

[12] Edwards (n 2), 19.

[13] Ibid, 37.

[14] Ibid, 37.

[15] Ibid, 46.

fashion,[16] or social vigilantism,[17] the pressure valve in the context of this thesis may operate more as a 'feature' of the system rather than a 'bug.' Such an approach may be applicable to the context of governing reuses of administrative data in research due to the nascent nature of many citizens' views and the variety and diversity of opinions.

In constructing the divergences suggested by the thesis' findings as a feature rather than a bug of the legal system, attempts to alter the law to reflect the views expressed within the empirical work are rejected. The empirical findings suggest such views could not be reflected due to their conflicting and often nascent nature. Such findings question how the law could demonstrate legitimacy in this context when the empirical evidence suggests that many individuals within society may find what is legally permissible to be inappropriate. How could it ever be found to be acting in the public interest where the construction of this concept remains contested? Furthermore, there is ambivalence as to what we mean by the 'public interest.' Gavaghan considers these questions[18] in the context of assisted reproductive technologies to ask whether the public interest should be defined by reference to the majority's views, or if certain views should be privileged, or if we should seek some common ethical denominators. The challenge is rendered starker by the comments made by Dickenson, who argues that attempts to uphold the public interest are 'intrinsically problematic in a liberal world-view, where

[16] Ibid.

[17] Ibid, 48.

[18] Colin Gavaghan, 'Reproductive Technologies and the Search for Regulatory Legitimacy' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017), 993.

there are no goods beyond the disparate aims of individuals, at best coinciding uneasily through the social contract.'[19]

Therefore, the challenges presented in this thesis are not entirely novel, as many acknowledge that innovations create challenges for law and regulation.[20] The thesis' findings agree with the argument made by Gavaghan that, in the context of regulating emerging technologies, the need for regulators to maintain a regulatory position that accords with a society's view of what is acceptable is an 'unusually demanding task.'[21] The challenges may not be new, but they highlight the 'importance of a space for contesting both complex knowledge claims, and what is deemed to be important about a decision.'[22] Furthermore, the fallacies of individuals attitudes towards risk are already well-acknowledged in the literature concerning behavioural law and economics.[23] This includes the 'availability heuristic' whereby individuals overestimate risks analogous to those covered by the media, 'hindsight bias,' where individuals overestimate the risks that have already materialised, and the 'status quo bias,' where individuals overvalue the benefit of preventing risks arising from new activities.[24] Within the focus group

[19] Donna Dickenson, 'The Common Good' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017), 142.

[20] Roger Brownsword, Eloise Scotford and Karen Yeung, *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017).

[21] Gavaghan (n 18), 992.

[22] Maria Lee, 'The Legal Institutionalization of Public Participation in the EU Governance of Technology' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017).

[23] Anthony Ogus, 'Economics and the Design of Regulatory Law' in Stephen Copp (ed), *The legal foundations of free markets* (Institute of Economic Affairs 2008).

[24] Ibid.

discussions, it is possible to discern some evidence of participants demonstrating the use of the availability heuristic and the status quo bias.

Considering this, the preferred response is to reject the notion that the law's enactment is the 'end point' upon which no further societal input is possible and is when the law's legitimacy must be judged. Rather than constructing the law as creating static institutional arrangements, understanding the law as fostering fluid regulatory spaces where a range of actors can participate in many ways is preferred. Doing so allows notions such as the 'public interest' to operate as a flexible vehicle open to development and evolution. Such evolution could occur through the continued interaction between lawmakers, those using the law to justify reusing data in research, and citizens. A level of flexibility is to be welcomed in this context. This allows for future changes to be based on further advances and the latest knowledge in this space.[25] Furthermore, this would capitalise on the divergences identified within this thesis as 'features' of the legal system, rather than 'bugs.' This approach chimes with the literature on the concept of a social licence for reusing data. A social licence considers 'how the expectations of society regarding some activities may go beyond compliance with the requirements of formal regulation.'[26] The problems that Carter and others identified in the context of reusing medical data in research, by defining this research as a 'fraught and contested activity, prone to concern over its ethical standing,'[27] appear to be

[25] Kevin Grecksch and Catharina Landström, 'Drought and Water Scarcity Management Policy in England and Wales—Current Failings and the Potential of Civic Innovation' (2021) 9 Frontiers in Environmental Science.

[26] Carter and others (n 5), 404.

[27] Ibid, 404.

reflected in the thesis' empirical findings. Therefore, their argument that there may be a need to 'go 'beyond compliance' with legal requirements' where activities are likely to 'provoke public disquiet,'[28] has relevance to this thesis.

The empirical findings support the argument that 'legal authority does not necessarily command social legitimacy.' [29] The findings thus align with the argument that a social licence may require 'something other than a legal mandate.' [30] Carter and others argue that this may be through the actors developing and delivering responsive risk-based approaches through mechanisms of principled proportionate governance.[31] This does not require a unanimous social consensus but rather that certain conditions of social engagement must be respected to foster genuine dialogic engagement.[32] This viewpoint resonates with the work of Aitken and others, who recognise that developing and maintaining a social licence requires the inclusion of diverse perspectives and interests.[33] As noted by Paprica and others, it is 'incumbent on the parties involved in research and data sharing to be transparent and to involve and engage with members of the public in an ongoing and authentic manner to ensure alignment with social

[28] Ibid, 405.

[29] Ibid, 408.

[30] Ibid.

[31] Ibid.

[32] Ibid.

[33] Mhairi Aitken and others, 'Establishing a Social Licence for Financial Technology: Reflections on the Role of the Private Sector in Pursuing Ethical Data Practices' (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720908892>.

licence.'[34] These ideas suggest that securing a social licence helps those using the law to justify their data sharing practices and demonstrate their legitimacy. In doing so, it helps establish the trustworthiness of the actors involved in reusing administrative data.[35] This aligns with Kerkin's writings on legitimacy, whereby goodwill and public confidence arise 'in part from institutions being given power in accordance with the constituted society's accepted norms and using it consistently with those norms.'[36]

Whilst the literature on social licences focuses on the effects of the actors seeking to secure the social licence, the interactions that help secure and maintain a social licence could also help secure the law's legitimacy. Beyond the impact upon the actors, it could enable key terms to incorporate societal views better. It could allow the law's flexibility, fostered through the language chosen, to balance the views of individuals in society whilst also recognising that a 'mirroring' of the law and societal views may not be feasible.[37] Here, the preferred approach is not to construct the social licence as completely distinct from the law. Instead, the conduct and approaches required by the social licence could develop an understanding of the key terms in law. This approach could then demonstrate the appropriateness of the legal approach towards reusing administrative data in

[34] PA Paprica, M Nunes de Melo and MJ Schull, 'Social Licence and the General Public's Attitudes toward Research Based on Linked Administrative Health Data: A Qualitative Study' (2019) 7 CMAJ Open E40.

[35] Sam HA Muller and others, 'The Social Licence for Data-Intensive Health Research: Towards Co-Creation, Public Value and Trust' (2021) 22 BMC Medical Ethics 110.

[36] Sarah Kerkin, 'Designing for Legitimacy: A Systems Perspective' (2017) 15 NZJPIL 67, 70.

[37] Tamanaha (n 8).

research. This entails a dynamic process that aligns with the writings on responsive law[38] as an ongoing process of norm development. The preferred approach is not to construct the social licence as parallel to the legal licence[39] but rather to feed into a developing understanding of the legal governance of reusing data in research. This reframes the approach away from seeking a final solution and to one of seeking improvements.

The understanding of legitimacy used in this thesis does not correspond to some Nonet and Selznick[40] discussions of the law's development. Nonet and Selznick argue the concept of 'legitimacy' is the 'central concern of autonomous law' where the main 'business of law is to certify the authority of rules and definitions.'[41] This understanding of legitimacy is tied to their view of the role of the court. In legal systems that comprise autonomous law, they argue that the system is court-centred and the expertise of the law 'lies in procedural fairness.'[42] Nevertheless, elsewhere they acknowledge that legitimacy 'must be seen as highly variable in its content and in effect.'[43] This diversity arises, in their view, as 'different principles differently affect the restraint of power and the opportunities for criticism of authority.'[44] It is acknowledged that procedural justice is necessary

[38] Nonet and Selznick (n 4).

[39] Nikki P Dumbrell, David Adamson and Sarah Ann Wheeler, 'Is Social Licence a Response to Government and Market Failures? Evidence from the Literature' (2020) 69 Resources Policy 101827.

[40] Nonet and Selznick (n 4), 104.

[41] Ibid.

[42] Ibid.

[43] Ibid, 55.

[44] Ibid, 56.

but insufficient alone, but legitimacy can incorporate aspects of both when determining the quality of the law. Therefore, a substantive understanding of legitimacy beyond formalistic tick-box approaches to incorporate substantive elements is proposed.

Based upon this understanding of legitimacy, the inclusion of citizens may enhance the legitimacy of laws by improving the law's quality through providing further information, access to dispersed knowledge, and a more deliberative collective problem-solving forum.[45] This aligns with their approach to regulatory legitimacy as a quality that may be 'enhanced by critical responses to regulatory concerns and consequent adaptation of structures and institutional environments when this is required to solve social problems.'[46] This view supports the increasing recognition that public disagreement is not a result of misunderstanding or irrationality and expert input is insufficient alone to resolve issues in a way deemed legitimate by society.[47]

This broadening of the actors to be included and the spaces in which they operate resonates with the term of governance. 'Governance' has a descriptive meaning of acknowledging the broader range of actors in different roles relevant to a system's governance through formal rule-making and more informal

[45] Lee (n 22), 622.

[46] Bärbel Dorbeck-Jung, 'Transcending the Myth of Law's Stifling Technological Innovation: How Adaptive Drug Licensing Processes Are Maintaining Legitimate Regulatory Connections' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017), 935.

[47] Lee (n 22).

interactions and negotiations.[48] Micheli and others also posit that governance has a normative scope that captures a way of rule-making that encompasses laws, regulations, standards, and other forms and guidance and the allocation of responsibilities and liabilities that promote wider participation.[49] In doing so, the participation of this broader set of stakeholders is acknowledged as a value of good governance[50] and could contribute to the law's legitimacy. This resonates with the arguments of Grecksch and Landström that normative, instrumental, and substantive grounds can justify the inclusion of a broader range of actors.[51] The normative rationale is rooted in the idea that people have a right to participate in matters that affect them. The instrumental rationale finds such participation to enable the effective implementation of decisions. The substantive rationale links public participation with higher quality decisions.[52]

This suggests that, for various reasons, citizens' greater inclusion in decision-making would be a positive development, but how could this be done? The following section explores the potential ways of operationalising this construction of the law and governance. It draws upon ideas such as the social licence and interactive legislation that create space for citizen involvement in determining what should be permissible when administrative data are reused in research.

---

[48] Marina Micheli and others, 'Emerging Models of Data Governance in the Age of Datafication' (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720948087>.

[49] Ibid.

[50] Ibid.

[51] Grecksch and Landström (n 25).

[52] Ibid.

9.3 OPERATIONALISING THE LAW

The thesis' key findings relate to the identification of divergences between the approaches of the GDPR and DEA towards governing the reuse of administrative data in social science research and individuals' views towards what is appropriate in this context. The incorporation of societal expectations, societal benefits, and the public interest within the legislative instruments and the documents that contributed to the legislative approach belies the diverse range of views demonstrated by the empirical findings. Where a socially constructed system of norms, beliefs, values, and definitions underpinned the law's legitimacy, this questions how far the law can be 'legitimate' where rapidly developing contexts challenges the ability to create such a system. However, these concepts remain central to justifying the legislative approach. Consequently, they provide space for data processors and controllers in this context to exercise their judgement, granting some flexibility that can adapt to new developments but also risks the embedding of divergences between what is legally permissible and societally acceptable.

This section considers how we can respond to these divergences, informed by the empirical findings that provided insights into why these divergences may have occurred. The findings support Lee's argument that

> *Law plays an important role in the institutionalization of public participation in the governance of new and emerging technologies. Law, however, also constrains participation, expressly and by implication restricting the range of matters that can be considered by*

decision makers and incentivizing particular approaches to explaining a decision.[53]

The previous section proposed a construction of legal legitimacy that depended upon ongoing interactions between lawmakers, those using the law to justify their actions, and citizens. The literature concerning social licence helped explore responses to the divergences identified. This could facilitate the law's development to further legitimise its approach by exposing citizens' norms, values, and beliefs that shape their views towards what is appropriate. It could ensure greater transparency, enhance accountability, and improve the representativeness of the decisions made. This resonates with shifts in approaches to governance within the EU. For example, Frahm and others have noted a move to 'inclusive, bottom-up strategies to foster innovation'[54] to respond, in part, to 'declining trust in institutions and a rise of desire for transparency and a greater degree of participation.'[55]

This section will consider the different ways this could occur, drawing upon the empirical findings and existing attempts to facilitate a greater role for citizens in determining what should be appropriate. The mechanisms' aims align with the adoption of a prudent regulatory stance. Dorbeck-Jung defines this as a commitment to four factors; enlarging 'the scope of inquiry to embrace knowledge of social contexts,' embracing 'a style of reasoning that reaches beyond mere rule-

[53] Lee (n 22), 620.

[54] Nina Frahm, Tess Doezema and Sebastian Pfotenhauer, 'Fixing Technology with Society: The Coproduction of Democratic Deficits and Responsible Innovation at the OECD and the European Commission' [2021] Science, Technology, & Human Values <https://journals.sagepub.com/doi/full/10.1177/0162243921999100>.

[55] Ibid.

following by focusing on social consequences,' 'testing alternative strategies for the implementation of mandates and reconstructing those mandates in the light of what is learned' and cooperating with all stakeholders and encouraging participation.'[56] This approach seeks to prevent these mechanisms from falling into the trap of providing a very procedurally oriented understanding of legitimacy. Doing so responds to the implied criticism by Nonet and Selznick regarding its procedural nature. This orientation risks being used to close down decision-making opportunities or legitimise pre-determined choices.[57] It also moves beyond the 'routine consultation provisions' that Lee criticises as a 'limited engagement with the practices of participation' due to the law's concern with 'individual rights than collective will formation.'[58] This does not detract from the importance of the processes used as they help provide a solid foundation for the meaningful involvement and engagement of a range of actors. Nevertheless, they are insufficient alone to secure the interactions proposed in this chapter.

As noted by van der Bung, a range of legal phenomena is emerging as part of a broader interactionist paradigm whereby lawmaking becomes a 'cooperative effort on the part of various stakeholders, of which the state is one, but not necessarily the most important.'[59] As part of this, the development of legal norms shifts from a vertical model to a 'more horizontal, interactive process in which

[56] Dorbeck-Jung (n 46), 935.

[57] Lee (n 22).

[58] Ibid, 626.

[59] Wibren van der Burg, 'The Emerging Interactionist Paradigm and the Ideals of Democracy and Rule of Law' in Bart van Klink, Britta van Beers and Lonneke Poort (eds), *Symbolic Legislation Theory and Developments in Biolaw* (Springer 2016), 37.

various social actors participate.'[60] This does not deny the instrumental function of legislation but emphasises its communicative and expressive functions whereby the fundamental values and aspirations norms are 'developed in a broad interactive process.' [61] This resonates with Fuller's writings, as explored in Chapter 5, which considers the law's legitimacy based on horizontal and vertical interactions and relationships.[62] Whilst some may find this provides an intolerable strain on the law by undermining its certainty and finality, the response of van der Burg notes that decisions by those using laws, such as courts, might be final, the legislation should remain open.[63] This approach would prevent a 'premature' closure that means the law soon requires changing or is ignored by those who are supposed to be guided by it in practice. Moreover, this approach may not undermine 'certainty' where this is defined as the ability of citizens to rely on reasonable expectations regarding others' behaviours as the implementation of laws arise from an open and inclusive process that they may be able to influence.[64]

A substantive understanding of legitimacy highlights that the mere use of these mechanisms will fail to enhance the law's legitimacy where it promises inclusivity that does not materialise or is inadequate. Such issues have been raised about the perceived failings of the 'GM Nation?'[65] debate undertaken by the UK

[60]Ibid, 42.

[61]Ibid, 43.

[62] Lon Fuller, *The Morality of Law.* (Yale University Press 1969).

[63] Van der Burg (n 59).

[64] Ibid.

[65] Alan Irwin, Torben Elgaard Jensen, and Kevin E Jones, 'The Good, the Bad and the Perfect: Criticizing Engagement Practice' (2012) 43 Social Studies of Science 118.

government in 2003 or the 2019 Convention Citoyenne pour le Climat.[66] A commitment to the law's legitimacy is not secured through the 'routine and mechanistic mobilisation of public engagement'[67] but rather through genuine attempts to better incorporate societal views as a highly relevant voice within the legislative process and post-legislative regulatory space. This commitment is more than the 'simple addition of values to objective expertise.'[68] Instead, the aim is to recognise citizens' 'intellectual substance'[69] through an active process that shapes the state's role[70] without romanticising certain forms of knowledge.[71]

This inclusion will not prevent uncertainty and indeterminacy but instead emphasises the importance of trustworthiness and the different framings of regulatory and legal choices.[72] A focus on trustworthiness, rather than trust, moves away from making citizens the 'problem.' The need to demonstrate trustworthiness places the onus on lawmakers, regulators, and those using the law to justify their decisions to inspire trust.[73] This chimes with Wynne's

[66] Vie Publique, 'Convention Citoyenne Pour Le Climat : Une Expérience Démocratique Inédite' (*Vie publique.fr,* May 18, 2021) <https://www.vie-publique.fr/eclairage/279701-convention-citoyenne-pour-le-climat-experience-democratique-inedite>.

[67] Frahm and others (n 54).

[68] Alan Irwin, 'STS Perspectives on Scientific Governance' in Edward J Hackett (ed), *The handbook of science and technology studies* (3rd edn., The MIT Press 2008).

[69] Brian Wynne, 'Creating Public Alienation: Expert Cultures of Risk and Ethics on GMOs' (2001) 10 Science as Culture 445.

[70] Irwin (n 68).

[71] Ibid.

[72] Ibid.

[73] Emily Rempel, Julie Barnett, and Hannah Durrant, 'Public Engagement with UK Government Data Science: Propositions from a Literature Review of Public Engagement on New Technologies' (2018) 35 Government Information Quarterly 569.

argument that issues of public understanding of science and public risk perceptions are 'not so much about public capabilities in understanding technical information, but about the trust and credibility they are prepared to invest in scientific spokespersons or institutions.'[74] To Wynne, the importance is not in the trust and credibility *per se* but the 'social relationships, networks and identities from which these are derived' that may be open to reconstruction through negotiation[75] and influenced by multifaceted networks. This broadening of the actors who merit inclusion moves the governance mechanisms away from what Irwin characterises as 'determinist ideas.'[76] Determinist ideas suggest that 'social controversies around science are 'really' all about politics or that complex areas of innovation can be reduced to 'social' construction' but that 'science and society are co-produced.'[77]

These mechanisms complement, rather than supplant, the traditional expressions of democracy, such as elections. Free elections are crucial to a democratic society, but a decision made by the ballot box may represent a citizen's view on many. Elections only occur every few years, potentially inhibiting citizens' potential to express their views at the ballot box towards rapidly developing technologies and their societal impact. Furthermore, these mechanisms may not prevent completely public backlashes or the negative media treatment of these developments. However, they provide a more productive space to respond to the

[74] Brian Wynne, 'Misunderstood Misunderstanding: Social Identities and Public Uptake of Science' (1992) 1 Public Understanding of Science 281, 282.

[75] Ibid.

[76] Irwin (n 68), 589.

[77] Ibid.

issues raised by these mechanisms and facilitate a more flourishing conception of democracy, which can reflect and respond to concerns raised by its citizens. The rest of this section considers potential governance mechanisms.

9.3.1 Lay Participation in Approvals Boards

Many different authorities and approvals bodies have used lay participation as part of their approvals process. It is used as a means of 'counteracting an exclusionary and socially disengaged policy tradition characterized by invocation of the objective authority of scientific expertise.'[78] Nevertheless, questions arise about the impact of a few people and their ability to act as a 'bridge' between 'scientific knowledge and broader publics.'[79] Such participation has, nevertheless, been used in approvals for medical research projects for decades. During the DEA's consultation process, respondents questioned why there was no opportunity for public participation in determining what was meant by the 'public interest' in research and the development of ethical projects.'[80]

ADR Scotland uses public panels as a 'growing element' of their public engagement work by 'providing a platform for members of the public to offer valuable feedback about our work.'[81] The aim is to 'test assumptions about public perceptions regarding administrative data research and involve the public in

[78] Irwin and others (n 65), 128.

[79] Ibid.

[80] Cabinet Office, 'Better Use of Data in Government Consultation Responses: Section 3' (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543829/better_use_of_data_consultation_response_section_3.pdf>.

[81] ADR UK, 'How Do We Work with the Public?' (ADR UK, 2020) <https://www.adruk.org/our-mission/working-with-the-public/>.

determining what 'public good' means in this context.'[82] However, one question that remains is how far this practice can contribute to a dialectical process within society and create feedback mechanisms between stakeholders and publics?[83] The lay participation tends to be confined to one individual being part of the panel, making it unclear how far this demonstrates the hearing of public voices rather than their marginalisation. Moreover, this mechanism may struggle to demonstrate its trustworthiness by failing to consider different histories and contexts of individuals and the complexity of public views on the government.[84]

9.3.2 Citizen Juries and Citizen Assemblies

Citizen juries and citizen assemblies are further means of viewing public engagement as a 'chain of crucial translations'[85] through lawmakers, data users, and citizens. Here, a larger public is translated into a smaller group of people who articulate their views on a matter that has been translated by a higher-level decision-making body into a matter to be put before the public.

This mechanism responds to one of the criticisms of lay participation in approval panels, where the potential for negotiating developments and creating feedback mechanisms is limited. Citizen assemblies and juries can incorporate more people and thus respond better to the façade of a monolithic society with one view towards appropriate reuses of administrative data in research. Nevertheless, there may be questions about the representativeness of the participants. Greater

[82] Ibid.

[83] Rempel and others (n 73).

[84] Ibid.

[85] Irwin and others (n 65), 126.

numbers do not automatically mean that certain sections of the impacted are not overlooked or that some may be 'seduced' in some way during the engagement event.[86] Furthermore, there may be difficulties framing a topic so that participants do not find the matter too complicated or superficial. This issue risks undermining participants' abilities to articulate in a 'proper and meaningful manner.'[87] There are also challenges in determining what a 'successful' engagement looks like in this context and how far these processes remain bounded by a technocratic ethos or a means to justify pre-determined choices.[88] Such issues, however, are not fatal to the use of citizen juries and assembles as serving to incorporate better the beliefs and opinions of participants towards what is appropriate in the context of reusing administrative data in research. Rather, they suggest that careful planning is critical and transparency regarding the process is critical to facilitating successful engagement.

Interestingly, some focus group participants described a mechanism akin to citizen assemblies as the desired means for ensuring appropriate data reuses. P4FG2 stated that

> *in an ideal world, you'd want like a jury kinda like setting with a diverse range of people which make up like the fabric society in this ideal society that you have them in government who decide public interest and that kind of thing … you kinda want that jury system of peers throughout society who can make, who can fairly make decisions based on all the groups.*

[86] Ibid, 126.

[87] Ibid.

[88] Ibid.

Others made similar appeals to more participatory forms of governance and deliberation. P2FG5 stated that they 'just think people should...kind of, guess, see what's being presented to the government somehow or be more involved in the process.' P2FG1 felt they would find secondary data uses more appropriate if 'had a different form of representative government in this country, maybe? errr if we could influence policy on more of a referendum style, like cantons in Switzerland or somewhere.'

The use of citizen juries and citizen assemblies raises the question of who is to oversee these forms of engagement and ensure they do not fall into the trap of merely demonstrating procedural legitimacy. Some participants considered this question within the focus groups, but there was no consensus on the answer. The discussions suggested a preference for an independent body to oversee this kind of mechanism. However, rather than discuss this further, the focus shifted to doubts about how 'independent' such a body would be. Such responses demonstrate the low trust in government as a key barrier to developing this form of participatory democracy, with P2FG2 stating that

> *it should be groups of individuals having processes in place to decide these things, but in general finding, like, an objective person or groups of people is always going to be very difficult so.... I think in my head I wouldn't want the government to be those people deciding what the public interest is…*

P4FG2 echoed this view, feeling that

> *ideally, you'd like to think you wouldn't need for it to be an external body, but probably would be best to have it external but yeah, nevertheless like diverse representatives throughout erm, but yeah, it's interesting, would probably lean to it being an external body.*

The potential candidates for this role could include the Information Commissioner's Office, the 'UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.'[89] Their guidance approach often emphasises the impact that data use may have on citizens and reflects efforts to encourage those using the law to meaningfully consider those affected by the decisions made by data users.[90] This is demonstrated in its collaboration with the Alan Turing Institute to conduct citizen juries. The aim was to understand citizens' views better and provide 'effective guidance to organisations about how to address data protection risks arising from technology.'[91] The UK government has also funded many initiatives that consider rapidly evolving technologies, including the Ada Lovelace Institute that already uses public engagement to inform their approaches.[92]

Previous citizens juries and assemblies have seen public bodies work with a range of partners to deliver them. For example, in 2020, Camden Council ran a citizens' assembly on health and social care to support policy development in this

[89] ICO, 'About the ICO' (2020) <https://ico.org.uk/about-the-ico/>.

[90] ICO, 'Guide to the UK General Data Protection Regulation (UK GDPR)' (2020) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>.

[91] ICO, 'Project ExplAIn Interim Report' (2019) <https://ico.org.uk/about-the-ico/research-and-reports/project-explain-interim-report/>, 5.

[92] See, for example, Ada Lovelace Institute, 'The Role of Good Governance and the Rule of Law in Building Public Trust in Data-Driven Responses to Public Health Emergencies' (2021) <https://www.adalovelaceinstitute.org/project/role-good-governance-rule-law-public-trust-data-driven-responses-public-health-emergencies/>.

area.[93] Alongside Camden Council, a consultancy acted as the independent lead facilitator, University College London provided academic input and an evaluation, and an expert advisory panel guided the approach. The expert advisory panel was also tasked with holding the public authorities to account for delivering the priorities that citizens discussed. The dissemination of a final report containing the citizens' proposed recommendations helped facilitate the delivery of the priorities discussed and enhanced accountability.[94]

In Ireland, the Irish Parliament established the Irish Citizens Assembly as an independent body with a Government-appointed Chairperson. [95] The Assemblies established in Ireland have considered topics such as abortion[96] and gender equality. [97] An external Expert Advisory Group, whose membership changes depending upon the issue to be considered by the Assembly, presents to or engages with Assembly members.[98] The Irish Government does not need to accept the recommendations contained within reports published by the Assembly, but it must respond to the legislature.[99]

---

[93] Participedia, 'Camden Health and Care Citizens' Assembly' (2020) <https://participedia.net/case/7429>.

[94] Camden Council, 'Camden Health and Care Citizens' Assembly' (2020) <https://www.camden.gov.uk/health-and-care-citizens-assembly>.

[95] Participedia, 'The Irish Citizens' Assembly' (2020) <https://participedia.net/case/5316>.

[96] Rónán Duffy, 'Analysis: Six Reasons Why Ireland Was a Landslide Yes for Repeal' *The Journal* (May 16, 2018).

[97] The Citizens' Assembly, 'About the Citizens' Assembly' (2019) <https://www.citizensassembly.ie/en/about-the-citizens-assembly/>.

[98] Ibid.

[99] Ibid.

These mechanisms enable a shift away from the façade of a universal understanding of the public interest. Taylor and Whitton find such a universal understanding problematic as it

> *is not self-evident within a pluralistic society which interests truly are common to all or even to most persons. Second, even if particular interests could be said to be common, this idea of the public interest does not explain how they should be traded in case of conflict.100*

Though contested, the concept of the public interest remains 'central' to the regulation of research.[101] Two responses in the literature shed light on how it can still provide insights into determining appropriate reuses of data in research. Sorbie argues in favour of a 'processual approach' to constructing the role of the public interest that is 'iterative and flexible, specifically drawing attention to the way that multiple actors, processes and interests interact, change and evolve over time.' This focuses on the dynamics of the legal and social frameworks within which decisions are iteratively constructed. This 'directs our attention to the dynamics of giving effect and legitimacy to law – not just on the page, but also in bringing context and meaning to its application both up and downstream.'[102] The dialogical approaches may not only be best suited to 'accommodating the conflicts, negotiations and adaptations inherent in the public interest' [103] but also

---

[100] Mark J Taylor and Tess Whitton, 'Public Interest, Health Research and Data Protection Law: Establishing a Legitimate Trade-off between Individual Control and Research Access to Health Data' (2020) 9 Laws 6, 16.

[101] Annie Sorbie, 'Sharing Confidential Health Data for Research Purposes in the UK: Where Are 'Publics' in the Public Interest?' (2019) 16 Evidence & Policy: A Journal of Research, Debate and Practice 249, 257.

[102] Ibid.

[103] Ibid, 261.

facilitating an understanding of what people will 'accept as justified in the circumstances.'[104]

To Taylor and Whitton, the focus on acceptability is an important consideration. They argue that the legitimacy of a trade-off is 'dependent upon public acceptance of its reasonableness: it is undermined if people do not accept it.'[105] In arguing in favour of such trade-offs being open to verification and that people 'must be able to see that the reasons for a particular trade-off are accessible and therefore open to challenge and correction.'[106] This openness may not be best facilitated through citizen juries and assembles, and larger forms of public participation and engagement may be necessary, an example of which follows in the next section.

9.3.3 vTaiwan

vTaiwan is a digital platform that provides Taiwanese citizens with the opportunity to discuss policies and produce consensus that can be used to create laws and policies.[107] The process consists of four stages, and a consensus from the participants drives transitions between the stages.

The first stage is the proposal stage, where issues can be submitted to a government authority that can accept or decline the proposal. This stage occurs every week through 'hackathons' where notes are taken using real-time

[104] Taylor and Whitton (n 100), 23.

[105] Ibid, 23.

[106] Ibid.

[107] Yu Tang Hsiao and others, 'VTaiwan: An Empirical Study of Open Consultation Process in Taiwan' (2018) <https://ideas.repec.org/p/osf/socarx/xyhft.html>.

collaborative tools, and information is shared through SlideShare[108] As of 2020, the vTaiwan mailing list includes 200,000 individuals.[109]

The second stage is the opinion stage, where different online platforms are used to gather opinions and enable responses. The platforms used during this stage include Discourse,[110] where the competent authority has seven days to reply to stakeholders' comments, and Pol.is,[111] which facilitates the engagement of large numbers of people by clustering people based on their responses to highlight the crucial aspects of the topic. Pol.is allows participants to post statements to which others can vote to agree, disagree, or pass and indicate if a statement is particularly important. The algorithm underpinning Pol.is then sorts people as a form of 'crowdsourced consensus mining' that makes the process of obtaining information about the 'dynamics of the issue, the facts of the matter, what is at stake, and who is involved' efficient and scalable.'[112] Reports are created based on the data that are publicly published, which highlight areas of consensus and polarising statements.[113]

The third stage is the reflection stage, which entails live-streamed online and offline facilitation sessions with stakeholders. During this stage, a summary of the process is given, and stakeholder groups can present their views. The

[108] SlideShare, 'SlideShare.net' (2020) <https://www.slideshare.net/>.

[109] Participedia, 'VTaiwan' (2021) <https://participedia.net/method/7387>.

[110] 'Discourse' (2021) <https://www.discourse.org/>.

[111] 'Pol.is' (2021) <https://pol.is/home>.

[112] CrowdLaw for Congress, 'VTaiwan' (2020) <https://congress.crowd.law/case-vtaiwan.html>.

[113] Participedia (n 109).

livestreams are monitored to allow for real-time responses to questions or requests for further information. The videos are subsequently uploaded online to provide all citizens with the opportunity to voice their views on the presentations.

The final stage is the legislation stage which presents the consensus reached from the previous stages. This consensus may not be at the level of detail of fully-fledged legislative provisions. Instead, the outcome may be a set of guidelines, a policy, or a draft bill. Nevertheless, the focus remains on determining what is acceptable to, rather than desired by, those involved in the process. vTaiwan has addressed over thirty issues, including Uber regulation [114] and policies on the sharing economy,[115] and the government has executed over 80% of the proposals.[116]

The process demonstrates a means by which citizens can remain engaged and bring their concerns to light in a forum where their views may make a difference. In seeking a rough consensus and allowing for a range of outcomes, it serves as an innovative, flexible, and useful tool through which the law can develop to new developments without alienating or angering citizens. The use of vTaiwan demonstrates that it is possible to use a constructive process by which different and diverse actors can collaborate in a meaningful way to contribute. This process

[114] Centre For Public Impact, 'Building Consensus and Compromise on Uber in Taiwan' (2019) <https://www.centreforpublicimpact.org/case-study/building-consensus-compromise-uber-taiwan>.

[115] vTaiwan, 'Sharing Economy' (2017) <https://vtaiwan.tw/topic/sharing-economy/>.

[116] Participedia (n 109).

is demonstrated through its use to determine the Taiwanese governance of UberX when it first entered the Taiwanese market.

9.3.3.1 vTaiwan and UberX

The introduction of Uber to Taiwan in 2013 prompted significant controversy. Whilst customers enjoyed cheaper fares and added convenience through the increased competition, existing taxi companies complained that Uber drivers were not subject to the same licensing and insurance requirements and nor did they pay the same taxes.[117] This was due to Uber registering in Taiwan as a technology company rather than a taxi firm.[118]

During the Opinion Stage, the vTaiwan community created an opinion survey on Pol.is whereby respondents could vote agree, disagree, or pass to statements and contribute their own. Nine hundred twenty-five respondents participated, providing 145 comments and 31,115 votes.[119] The algorithm underpinning Pol.is identified four broad opinion groups comprising taxi drivers, UberX drivers, UberX passengers and non-UberX passengers. These were merged into two groups that disagreed over whether the introduction of UberX had been a positive development. The survey did identify a strong consensus on the statement that

> *The government should leverage this opportunity to challenge the taxi industry to improve their management & quality control system,*

[117] Carl Miller, 'Crossing Divides: How a Social Network Could Save Democracy from Deadlock' *BBC News* (October 25, 2019.

[118] Yu Tang Hsiao and others (n 107), 3.

[119] Ibid, 4.

so that drivers & riders would enjoy the same quality service as UberX's[120]

In response, vTaiwan enabled the groups affected by the introduction of Uber to be involved in the government's potential regulatory response. In total, 1875 online participants joined a live-streamed meeting with academics, industry experts, and representatives from the Association of Taxi Drivers in Taipei, Taxi Taiwan, Uber, and the Ministries of Transport, Economic Affairs, and Finance.[121] The stakeholders agreed that Uber should comply with the law regarding insurance and driver training.[122] The final stage saw the administration pledge to ratify the consensus identified through the Pol.is survey through amending the Regulation on Automobile Transportation Management.[123]

9.3.3.2 Opportunities and Constraints

As noted in this chapter, mechanisms are not a silver bullet to resolving the issues raised by divergences uncovered in this thesis. Divergences may not always require a change in the law to align better with societal expectations, but they highlight a need to engage meaningfully with the public. By involving a greater range of people in the governance of reusing administrative data in research, it can help create the

[120] Ibid, 4.

[121] Audrey Tang 唐鳳, 'Uber Responds to VTaiwan's Coherent Blended Volition' *Medium* (May 23, 2016).

[122] Yu Tang Hsiao and others (n 107), 4.

[123] Regulation on Automobile Transportation Management. <http://law.moj.gov.tw/LawClass/LawContent.aspx?PCODE=K0040003>.

> *checks and balances that engender greater societal and economic equity, can help to rebalance asymmetries of power, and can contribute towards increased public confidence in the use of data.*[124]

The Ada Lovelace Institute finds that effective participatory approaches can generate a 'beneficial feedback loop.' Therefore, participatory approaches can create a strong sense of trust and confidence that leads to more equitable outcomes, increasing the social licence in this context.[125] Nevertheless, as noted by Veress, 'transparency and mutual trust were preconditions and outcomes of cooperation.'[126] The empirical findings cast doubt upon whether there is the sufficient mutual trust required to foster the kind of collaboration that could respond to the divergences identified. As noted by the Centre for Data Ethics and Innovation, the current trust of the public in sharing data by the public sectors is 'tenuous.'[127] The environment of 'tenuous trust' results in communication to the public that is 'primarily to limit a potential negative reaction, rather than active, positive engagement.'[128] The Centre for Data Ethics and Innovation suggested that a lack of understanding of what might be 'publicly acceptable' in the context of data sharing by public bodies may 'create perverse incentives for departments to be less transparent about their work.'[129] This risks revelations about secondary

---

[124] Ada Lovelace Institute, 'Participatory Data Stewardship' (2021) <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/>, 10.

[125] Ibid, 11.

[126] Jozsef Veress, 'Modelling Civil Society's Transformational Dynamism and Its Potential Effects,' *European Conference on Modelling and Simulation* (2017) <http://www.scs-europe.net/dlib/2017/ecms2017acceptedpapers/0106-fes_ECMS2017_0096.pdf>, 5.

[127] Centre for Data Ethics and Innovation, 'Addressing Trust in Public Sector Data Use' (2020) <https://www.gov.uk/government/publications/cdei-publishes-its-first-report-on-public-sector-data-sharing/addressing-trust-in-public-sector-data-use>.

[128] Ibid.

[129] Ibid.

uses of data may 'come as a surprise' and in the event of something going wrong 'the public reaction may be particularly difficult to manage.'[130]

Furthermore, whilst much of the vTaiwan process may be transparent in the sense of the information shared throughout the process and the use of open-source platforms,[131] the Pol.is aspect of the process provides some challenges. Firstly, the modelling of opinions is 'reliant on subjective assumptions about what content is relevant,' which is amplified by reducing the information into clusters of opinion groups.[132] Secondly, the algorithm that determines the clusters may create issues of explainability.[133] However, the UK government found the focus on 'consensus' constructive.[134] The ability to vote on statements rather than respond to them inhibits trolling, preventing the process from being side-tracked.[135]

The mechanism of vTaiwan, which may be seen as a developed example of 'CrowdLaw,'[136] better enables more expertise to be brought to bear when considering law and policy responses to complex problems through collaborative approaches. This is particularly the case with vTaiwan as it is an institutionalised

[130] Ibid.

[131] Simon Vydra and others, 'Big Data Ethics: A Life Cycle Perspective' (2021) 14 Erasmus Law Review.

[132] Ibid.

[133] Ibid.

[134] Department for Business, Energy & Industrial Strategy, 'The Use of Public Engagement for Technological Innovation Literature Review and Case Studies' (2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955880/use-of-public-engagement-for-technological-innovation.pdf>, 129.

[135] Ibid.

[136] Beth Simone Noveck, 'Crowdlaw: Collective Intelligence and Lawmaking' (2018) 40 Analyse & Kritik 359.

process that requires a response from the national government.[137] Nevertheless, as an institutionalised process, it can struggle to 'retain authority in the selection of topics' to be considered and cannot proceed without agreement from a government authority.[138] vTaiwan also aligns with the idea of 'civic tech,' defined by Russon Gilman and Carneiro Peixoto as 'technology that is explicitly designed and leveraged to increase and deepen democratic participation.'[139] The primary concern is to expand the relationship between the citizen and the state by enhancing modernisation and efficiency and offering 'further channels for citizen feedback, voice, accountability, and governance transparency.'[140] vTaiwan can lower the barriers to entry. However, there is a risk of some exclusion by relying on digital tools that may bias its use towards more tech-confident citizens. The broader literature on mechanisms of participatory democracy recognises this risk, finding that 'participation is skewed to those who are already politically active and towards well-educated, young men in urban areas.'[141]

This suggests that there is a need to explore further the ability to evaluate such mechanisms to determine how far they improve the quality of decision-making by legislators and the quality of the relationships between legislators, citizens, and those using the law. Whilst the anecdotal evidence appears positive,

---

[137] Ibid, 368.

[138] Yu Tang Hsiao and others (n 107), 5.

[139] Hollie Russon Gilman and Tiago Carneiro Peixoto, 'Digital Participation' in Stephen Elstub and Oliver Escobar (eds), *Handbook of Democratic Innovation and Governance* (Edward Elgar Publishing 2019), 106.

[140] Ibid.

[141] Julie Simon, Theo Bass and Victoria Boelman, 'Digital Democracy: The Tools Transforming Political Engagement' (2017) <https://www.issuelab.org/resources/33129/33129.pdf>, 83.

more robust forms of analysis would be helpful to demonstrate the potential impact of such mechanisms.[142] This evaluation could consider the 'input legitimacy' of outcomes, whereby the result is responsive to the will of the people, the 'output legitimacy,' which considers how far the actions taken to provide solutions, and the 'process legitimacy,' which concerns the fairness, transparency, and inclusivity of the process.[143] The analysis of these different types of legitimacy will help deepen the understanding of the relationship between them. Lindgren and Persson[144] have noted considerable disagreement in the literature as to whether input and output legitimacy are mutually reinforcing or may result in a trade-off between them. Their empirical study has suggested a synergy between these forms of legitimacy but the authors acknowledge that future research would be beneficial to explore this synergy further.[145] Furthermore, Deligiaouri and Suiter[146] have suggested that these forms of legitimacy should be examined by an overall criterion of 'functional legitimacy' that may justify the selection of different inputs and outputs to best facilitate the legislative process by emphasising the need for a well-thought strategy.

Based upon the challenges presented in these options, the final mechanism explored in this chapter is the creation of a public data trust. This mechanism

---

[142] Ibid, 84.

[143] Ibid, 85.

[144] Karl-Oskar Lindgren and Thomas Persson, 'Input and Output Legitimacy: Synergy or Trade-Off? Empirical Evidence from an EU Survey' (2010) 17 Journal of European Public Policy 449.

[145] Ibid.

[146] Anastasia Deligiaouri and Jane Suiter, 'Evaluation of Public Consultations and Citizens' Participation in 2015 Better Regulation Agenda of the EU and the Need for a Deliberative E-Rulemaking Initiative in the EU' (2020) 22 European Politics and Society 69.

shifts away from the explicit focus on including individuals present in the other mechanisms, though it is not without its own challenges.

9.3.4 Public Data Trusts

Though presented as a separate mechanism, establishing a public data trust could be supported by some of the mechanisms already discussed within this chapter. The relationships that may be forged between these mechanisms serve to reinforce the point that a commitment to enhancing the law's legitimacy should not be confined to one-off tick-box exercises but rather an ongoing participatory process that uses a range of tools to fulfil its aims.

Public data trusts refer to a data governance model that allows for a public actor to access, link, and use data on citizens.[147] The data stored within a public data trust may also include that created and held by commercial entities, which may be useful in the context of modern governance within the UK where there is an increasing reliance on commercial entities to provide services often associated with public bodies.[148] Public actors become trustees of the data, and their role is to ensure that citizens' data are handled ethically, privately and securely. The underlying principle of a public data trust is that 'all data with a public interest component is part of a national infrastructure, and therefore the information it

[147] Micheli and others (n 48), 8.

[148] Peter Blackburn, 'Outsourced and Undermined: The COVID-19 Windfall for Private Providers' *The British Medical Association* (September 8, 2020) <https://www.bma.org.uk/news-and-opinion/outsourced-and-undermined-the-covid-19-windfall-for-private-providers>.

affords should be 'socialised' to produce value for 'citizens and society as a whole.'[149]

The relationship generated from a public data trust implies a level of trust between the public actors, as the trustees, and the citizens, who are the beneficiaries of the public data trust. Citizens are categorised as beneficiaries of such trusts through the insights that can be gained from their data and their implementation to create policies.[150] The empirical findings of this thesis cast doubt as to whether such a relationship of trust exists. However, the use of citizens assemblies or a process similar to vTaiwan could potentially assist in fostering such trust.

A public data trust can also incorporate external independent organisations to act as trusted intermediaries. These trusted intermediaries seek to ensure that data is securely managed and used in a way to maximise its value for citizens.[151] As noted previously, the empirical work of this thesis also cast doubt on the potential for independent organisations to foster trust in this context. The findings suggested that certain participants mistrusted the 'independent' status of certain bodies that operated to hold the government to account. Furthermore, some participants felt that many bodies, including charities, were more concerned with pursuing profits or their own objectives rather than the public interest.

[149] Micheli and others (n 48), 9.

[150] Andrew Collinge, 'A 'New Deal' for City Data? – London Datastore' (2018) <https://data.london.gov.uk/blog/a-new-deal-for-city-data/>.

[151] Micheli and others (n 48).

9.4 CONCLUDING THOUGHTS AND REFLECTIONS

How does the law approach the governance of reusing administrative data in social science research, an area that has rapidly developed due to technological advances that enable greater insights but foster greater risks towards citizens and their interests? This thesis explored the governance of this data processing activity, testing the claims that citizens' views are incorporated when determining what should be legally permissible. A socio-legal approach aided this thesis' exploration of the research questions, contributing to the literature on the interplay between the law and citizens within a society. The previous empirical work focuses upon participants views without relating them to the legal governance approach. Conversely, the literature exploring what is legally permissible in this context has tended not to consider societal views towards what has been rendered lawful. Some of the literature on care.data has adopted a socio-legal approach,[152] and this thesis has explored this within the oft-overlooked context of reusing administrative data in research. The thesis demonstrated that this interplay between the law and societal views is important and considered previous work in a different setting.

This thesis has contributed to the existing literature by showing how the legislative approaches have come to be through a detailed account of the forces that shaped the law. The empirical work has shed light on the law's reception within society and views towards how far the law aligns with individuals'

[152] See, for example, Carter and others (n 5) and Paraskevas Vezyridis and Stephen Timmons, 'Understanding the Care.data Conundrum: New Information Flows for Economic Growth' (2017) 4 Big Data & Society 1.

understandings of what is appropriate in this context. This allowed for a reflection upon the legitimacy of the law, suggesting reuses of administrative data in research that were lawful yet inappropriate to many participants and potential causes of this. It demonstrated that legitimacy, understood in these terms, would be difficult for the law to attain at the point of its passing as the rapidly developing nature of data processing and analysis and the remoteness of reusing data undermined the ability to speak of a single system of socially constructed norms.

This final chapter considered potential responses to the problems identified within the thesis. It sought to redefine legitimacy not as an attribute of the law ascribed at its passing but as an ideal to strive for continually. In premising legitimacy upon socially constructed views and beliefs, this created space for the ex-post involvement of citizens to help further shape the law. The literature on securing a social licence illuminated how understandings of what is appropriate could be developed. I argued in favour of using this licence to shape and develop the law rather than see it as a parallel mechanism for governance. In doing so, I outlined mechanisms that could foster greater public engagement and support the overall aim of enhancing the law's legitimacy in this context. There is no silver bullet offered in this thesis but a focus on substantive involvement, an acceptance of the pluralistic nature of modern society and its impact on our understanding of the 'public interest,' and a shift away from what is preferred to what is acceptable to citizens are critical aspects of such governance.

9.4.1 Future Directions

The thesis suggests several productive avenues for future research. I chose to focus on the practices that shaped the legislative outcomes of the GDPR and DEA but

left unexplored how these are used in practice. To better understand the effects of these legislative instruments, there is a need to understand how actors, including researchers and officials within the government, have responded to them. This thesis has provided empirical evidence to support the claim that what is legally permissible may diverge from many individuals' views on what are appropriate reuses of administrative data in research. However, it remains unclear how these divergences may already be negotiated by those who use the law to justify their data processing activities. This would require further empirical work to explore the effects of the legislation on these actors' determination of what is permissible. There is considerable scope for these insights to contribute to our understanding of what can be done where the law diverges from what its citizens deem to be appropriate uses of data.

This thesis provided interesting insights into participants' views towards reusing data in research through a mixed-methods approach. These insights enabled the exploration of why participants may hold such views. The findings suggest that participants' views are complex, drawing upon a range of personal experiences to provide context-sensitive accounts as to what is appropriate. Further empirical research could shed further light on these views and their potential role in the post-legislative space. Doing so could provide insights into which mechanisms could facilitate the meaningful post-legislative engagement, enabling the relationship between the law and its citizens to flourish as they respond and adapt to changes in the data processing space. This thesis has provided the first step towards potentially reconfiguring this relationship and the

role of the state and citizens in constructing concepts and categories that shape how we respond to reuses of administrative data in social science research.

ANNEX A

This Annex balances the need to provide a detailed explanation of the methods and justification of the methodological choices whilst retaining a strong and coherent narrative throughout this thesis. Therefore, Chapter 1 outlines the empirical approach. Chapter 4 justifies using a mixed-methods approach based upon the research questions and the importance of societal views. The Annex provides the research protocol and the responses necessitated by the pandemic. The Annex commences with justifying the mixed methods approach before detailing the quantitative and qualitative research that comprises the empirical work.

MIXED METHODS

This research uses a mixed-methods approach, combining an online survey and a series of online focus groups to respond to the research questions explored within this thesis. Specifically, the approach is an explanatory sequential mixed methods approach, where the qualitative research follows the quantitative research. This consists of first collecting quantitative data and then collecting qualitative data to help explain or elaborate on the quantitative results.[1]. The quantitative data suggested areas of divergence between the legislative approaches of the GDPR and the DEA and respondents' views towards the appropriateness of reusing data in research. Nevertheless, the quantitative analysis suggested inherent limitations of a quantitative approach in responding to the research questions. The qualitative

[1] John Creswell and Vicki Plano Clark, *Designing and Conducting Mixed Methods Research* (Sage 2018)

research sought to counter these limitations. The qualitative research was further informed by the legislative analysis in Part I and existing research in this area. However, it still sought to include some freedom for participants to make novel contributions. This would all the qualitative data to help refine or extend the picture provided by the quantitative data.

Method choices have implications as the methodologies they are part of have a specific theoretical perspective. These perspectives relate to the ontological underpinnings, about the nature of existence, and epistemological underpinnings, regarding the nature of knowledge. Qualitative research is associated with subjectivism, and quantitative research aligns with positivism, leading some to argue that they are incompatible. The positivist ontology proposes a single reality, whereas a constructivist ontology sees the nature of reality as multiple. The epistemologies of these two 'worldviews'[2] also differ, whereby positivism seeks objective knowledge, and subjectivism utilises a more subjective understanding of knowledge, which comes from individuals' experiences and perspectives. Creswell and Piano Clark[3] argue that the research design can justify using multiple paradigms. The underlying premise is that the guiding assumptions of the worldviews shape the construction of the research procedures. For example, aspects of a post-positivism worldview may be identified in quantitative research. The survey encompassed specific variables and was framed towards testing the theory that the legal approach and societal expectations towards reusing

[2] Ibid.

[3] Ibid.

administrative data in social science research diverge. The worldview then shifts in the qualitative phase of the empirical work. The aim was to enrich our understanding and explore why these divergences may occur.

The views of Creswell and Piano Clark align with the work of Feilzer.[4] Feilzer argues that positivism and subjectivism research paradigms both seek to find the truth, whether this is an objective one or a relative truth of multiple realities. Therefore, both paradigms attempt to produce knowledge that best represents reality. These commonalities have led many to question the need for these divisions as a meaningful way of understanding the purpose of research.[5] This suggests that methodological positions are perhaps less distinct than previous literature suggested. Dewey's notion of the world as an 'existential reality' of different layers leads Feilzer to argue persuasively that mixed methods explore these layers.[6] Therefore, the constructivist underpinnings of Part I of this thesis are not necessarily incompatible with the quantitative aspect of the empirical research.

PLANNING

The planned timeframe for the initial empirical work coincided with the lockdown and consequent restrictions on movement, requiring a significant shift in the empirical approach. The initial plan was to survey people face-to-face, travelling

[4] Martina Yvonne Feilzer, 'Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm' (2009) 4 Journal of Mixed Methods Research 6.

[5] R Burke Johnson, Anthony J Onwuegbuzie and Lisa A Turner, 'Toward a Definition of Mixed Methods Research' (2007) 1 Journal of Mixed Methods Research 112, 117 and Feilzer (n 4), 8.

[6] Feilzer (n 4).

to busy places, including shopping centres and libraries. The plan was to create a 'data donation card,' allowing respondents to permit or deny the donation of aspects of their data to certain entities for specific purposes. As the surveys would have been paper-based and taken in my presence, I would have encouraged participants to write further information on these surveys. Respondents could explain or justify their views or highlight things of importance and relevance to their responses that were not included on the card. I aimed to use this data donation card and my interactions with participants to gain a deeper insight into participants' views than a traditional survey would allow. This became impossible and, to enable some data collection to take place, I decided to undertake a more traditional survey. Therefore, the initial lockdown phase prompted a significant alteration of the empirical approach to respond best to the limitations whilst still exploring the research questions.

The ongoing restrictions led to using online focus groups rather than hosting them in community centres, libraries, and other shared spaces. This decision was taken after a consideration of the issues regarding their exclusivity due to the digital divide.[7] Nevertheless, online focus groups were the only practicable means to explore participants' opinions, justifications, and experiences. Furthermore, using online methods allows individuals to participate without travelling to a central venue, allowing those in isolated areas or with accessibility needs to participate. Research also suggests that using online

[7] Carl Baker and others, 'COVID-19 and the Digital Divide' (*POST,* December 17, 2020) <https://post.parliament.uk/covid-19-and-the-digital-divide/>.

methods need not hinder the richness of data collected.[8] There are many similarities between conventional and online focus groups, such as the real-time nature of the exchange, the use of verbal and non-verbal cues, and enabling personable interactions that may demonstrate spontaneity.[9] Some accounts have suggested that online focus groups may foster more detailed and conversational by enabling a more symmetrical relationship between the research and the researchers.[10]

THE QUANTITATIVE METHODOLOGY

Survey Design

The survey questions were based upon two sources of information. The first source of information was the existing legislation that governs reuses of administrative data for research purposes. The legal analysis in Part I suggested factors in the legislative instruments that determined permissible and impermissible data uses, such as the identifiability of data. Secondly, previous empirical work that examined participants' attitudes towards reuses of data influenced the empirical work.[11] These suggested factors that were not relevant to the legal governance of secondary uses of administrative data in research but were relevant to shaping respondents' views. These sources of information were used to explore the

[8] Nigel Fielding, *The SAGE Handbook of Online Research Methods* (Sage 2017).

[9] Marnie Howlett, 'Looking at the 'Field' through a Zoom Lens: Methodological Reflections on Conducting Online Research during a Global Pandemic' [2021] Qualitative Research <https://journals.sagepub.com/doi/epub/10.1177/146879412098569>.

[10] Ibid.

[11] This literature is explored in Chapter 6, which presents the quantitative findings.

potential existence of divergences between the legal approaches to the appropriateness of reusing administrative data in social science research and respondents' views.

The survey was created using the survey builder available through Jisc online surveys,[12] which the University of Oxford recommends this site as it is GDPR-compliant and accords with the information standards required by ISO27001. The survey received approval from the Social Sciences & Humanities Interdivisional Research Ethics Committee,[13] one of three subcommittees of the Central University Research Ethics Committee (CUREC).[14] The survey was designed to take approximately 15 minutes to complete, and respondents were incentivised to participate through a prize draw for a £20 shopping voucher.[15] Respondents could complete the survey without submitting an email address, but the lack of contact details meant they could not enter the prize draw. A separate question asked if respondents' email addresses could be used to contact them about future empirical work. This ensured that this consent was free and fully informed as it remained independent from the prize draw. The full survey is provided in Annex B.

[12] Jisc, 'Online Surveys' (2021) <https://www.onlinesurveys.ac.uk>.

[13] University of Oxford, 'Social Sciences & Humanities Interdivisional Research Ethics Committee (SSH IDREC)' (2021) <https://researchsupport.admin.ox.ac.uk/governance/ethics/committees/sshidrec>.

[14] University of Oxford, 'Research Ethics (Including CUREC)' (2021) <https://researchsupport.admin.ox.ac.uk/governance/ethics>.

[15] Of the participants who expressly gave their email address to allow themselves to be entered into the prize draw.

Overall, 50 survey items sought to assess respondents' views towards the acceptability of the type of administrative data that could be reused in research, the identifiability of the data reused in research, the actors involved in this data processing, the purpose of the data processing, and the ability to link data. Each item had a five-fold Likert scale; very unacceptable, quite unacceptable, neither acceptable not unacceptable, quite acceptable, and very acceptable. For two questions, respondents were also given the option of selecting 'not applicable' as it was recognised that not all items may be relevant to their situation. For example, some selected religious data and data on the use of Council-owned facilities as not applicable. A Likert-type scale was used to measure respondents' views as it is simple to use, providing respondents with choices without overwhelming them. Furthermore, it provides an option for a neutral response to prevent the distortion of result. Sociodemographic information was also collected from respondents pertaining to age, gender, ethnicity, and employment status. These characteristics were collected through categorical variables to help preserve respondents' privacy.

While a survey comprising of open questions may have gained more data, the longer time frame may have deterred respondents. The thesis explores a topic that individuals rarely consider. Therefore, many do not have a pre-existing 'stake' in the matter that can encourage people's participation. The survey did incorporate two open-ended questions. These questions provided data on respondents' views towards different organisations accessing their data and the potential for data linkage. Open questions would allow participants to explain their reasoning further or suggest something not included in the original survey. This collected a small amount of qualitative data without entailing an onerous

time burden on respondents. Of the 302 participants who completed the survey, 60 commented on which organisations they would not want to access their data and 40 shared their views on data linkage.

Survey Reliability and Validity

The survey was piloted on a small number of individuals (n=6) to ensure its legibility and accessibility. The feedback prompted a change to the question order. The survey started with the question proving respondents with different administrative data types. This question order helped participants to understand better the breadth of administrative data to inform their responses. Cronbach's alpha was used to assess the internal consistency of the survey. A result of $\alpha = .969$ suggests a high level of internal consistency within the items.[16]

Survey Recruitment

The survey was open to all people over the age of eighteen and living in the UK. Recruitment took place through several means including the sharing of the survey with my personal networks who then shared it with their networks. This is a method of convenience sampling as it cannot provide a random sample but the time and financial cost of collecting a random sample meant this was not possible within the bounds of the DPhil. This effect is potentially reinforced by the lockdown restrictions necessitating the use of online methods only to disseminate the survey, which may introduce a level of bias as those with no or low digital literacy may have been excluded. The survey was also published on Call for

[16] Andy Field, *Discovering Statistics Using SPSS (and Sex and Drugs and Rock 'N' Roll)* (Sage 2012)

Participants,[17] which, for a small fee, creates a webpage with an overview of the study, a direct link to the survey, a poster for distribution, and links to share via various social networks. The website also sends a copy to those who have signed up to its mailing list. Again, this is a form of convenience sampling but enabled the quick and cost-efficient recruitment of respondents. Finally, the survey was shared on numerous community Facebook groups, and efforts were made to distribute the survey in groups linked to different locations, such as Community Centres in Wales, and Churches in Scotland.

The survey opened on 19th October and closed on 16th December 2020, after receiving 302 completed responses. Initially, the survey was programmed to close after 300 responses, but I decided to re-open it to capitalise on the loosening of restrictions that allowed community centres to partially re-open. The intention was to include more responses from those who could access the Internet, and the survey, through community-based facilities. Two further responses were collected, but it was unclear whether this was from community centre users as many may not have reverted to their usual habits after the lockdowns.

Respondent Demographics.

The survey received 302 complete responses. Table 1 provides information on the respondents' ages. The use of categorical variables helped further protect participants' privacy and encourage them to provide a response. Most respondents were aged between 25 and 30 and over sixty two percent were aged 40 or under.

[17] Call for Participants, 'Call for Participants' (2021) <https://www.callforparticipants.com>.

This indicates that there may be limitations to the survey's findings as the data is more heavily weighted to younger people.

Table A.1: Age Range of Survey Respondents

Age	Frequency	Percentage
18 – 24	66	21.9
25 – 30	71	23.5
31 – 40	52	17.2
41 – 50	32	10.6
51 – 60	45	14.9
61 – 70	17	5.6
Over 70	12	4.0
Prefer not to say	7	2.3
Total	**302**	**100.0**

Table 2 provides the respondents' genders. Within the study 72.2% (n=218) identified as female, 23.8% (n=72) identified as male, 1.3% (n=4) identified as non-binary, 0.3% (n=1) identified as other, and 2.3% (n=7) preferred to not give a response. It is acknowledged that such a skew towards those identifying as female means these findings do not reflect the general population.

Table A.2: Gender of Survey Respondents

Gender	Frequency	Percentage
Male	72	23.8
Female	218	72.2
Non-binary	4	1.3
Other	1	.3
Prefer not to say	7	2.3
Total	**302**	**100.0**

Table 3 provides information on the respondents' ethnicities. When asked about their ethnicity, 81.7% (n=247) identified as White British or White Other, 13.9% (n=42) identified as being from a BAME background, and 4.3% (n=13) preferred not to give their ethnicity. This recoding of the categories into smaller

categorical variables was done to the low numbers of responses in some categories, as demonstrated in Table 3, which can affect impact upon the ability to statistically analyse the data. The ethnicity items were based upon the options given in the UK's Census.

Table A.3: Ethnicity of Survey Respondents

Ethnicity	Frequency	Percentage
White British	207	68.5
White Other	40	13.2
Asian	22	7.3
African	3	1.0
Caribbean	2	.7
Arab	1	.3
Mixed	7	2.3
Other	7	2.3
Prefer not to say	13	4.3
Total	**302**	**100.0**

Finally, respondents were asked about their employment status, which is shown in Table 4. Most respondents were in employment (n=174, 57.6%) followed by being in education (n=76, 25.4%)

Table A.4: Employment Status of Survey Respondents

Employment Status	Frequency	Percentage
Full-time work	135	44.7
Part-time work	39	12.9
Unemployed	7	2.3
Retired	20	6.6
Student	76	25.2
Other	17	5.6
Prefer not to say	8	2.6
Total	**302**	**100.0**

Analysis

IBM SPSS Statistics 27[18], a software platform for statistical and textual analysis, analysed the survey data. The software is often used for the data analysis discussed in this thesis and allows data to be password protected, a condition of the CUREC approval. Descriptive statistics were calculated as frequencies and percentages to summarise the variables of interest. Bar charts were created to display the distribution of responses to facilitate a comparison between respondents' views and the legislative approaches analysed in Part I to test the hypothesis that these approaches would diverge. As ordinal data was collected via Likert-type scales, non-parametric tests were performed on the data to explore the existence of statistically significant divergences between the respondents' views and the legislative approaches, including Mann Whitney tests and Kruskal-Wallis tests.

THE QUALITATIVE METHODOLOGY

Justifying the Use of Online Focus Groups.

As noted in Chapter 7, focus groups are an apt method for providing insights into what participants think and the drivers and justifications that prompt their views. An internship that I undertook at Policy Lab,[19] which advocates a human-centred approach to policymaking, demonstrated the strengths of qualitative research. I learned about the potential for lay expertise to contribute to the law and

[18] IBM, 'SPSS Software' <https://www.ibm.com/analytics/spss-statistics-software>.

[19] Policy Lab, 'About Policy Lab' (2014) <https://openpolicy.blog.gov.uk/about/>.

policymaking processes and the benefits of including the public in decision making.

I considered alternatives to focus groups during the planning of the empirical work. These included hosting citizen assemblies[20] or citizen juries[21] to collect further data on participants' attitudes towards reusing data in research. Unfortunately, my research on these methods emphasised the need for substantial amounts of time, money, and personnel to facilitate them properly. Furthermore, the restrictions imposed in response to the pandemic suggested my empirical research would have to be online. Citizen juries and citizen assemblies tend to last over a couple of days.[22] Keeping participants engaged for this length of time through online means would have been incredibly difficult. The lockdown restrictions and the constraints imposed by the project in terms of available time and financial resources meant online focus groups were the most suitable method.

Planning

The planning of the focus groups entailed making various decisions, including the choice to host the focus groups online. Other decisions included the questioning approach, the recruitment protocol, and the practical aspects of the focus groups. The data required to explore the research questions, the time and financial limits

[20] Participedia, 'Citizens' Assembly' (2020) <https://participedia.net/method/4258>.

[21] Participedia, 'Citizens' Jury' (2018) <https://participedia.net/method/155>.

[22] See, for example, Participedia, 'Citizens' Assembly on Brexit' (2017) <https://participedia.net/case/5166>.

of a DPhil thesis, and the points raised by the ethics committee informed these decisions.

CUREC,[23] the University of Oxford's central research ethics committee, approved the empirical research. This approval encompassed the decision to use Zoom Cloud Meetings,[24] the recruitment process, the questioning route, and the safeguards to protect participants' interests best. Using Zoom required further justification from the research ethics committee, who were concerned about the safeguards the platform uses to protect individuals' data and proposed using Microsoft Teams. I felt that Microsoft Teams, which requires the purchase and installation of expensive software, could create barriers to participation. I also felt Zoom was a more familiar platform to many, and its use spans both professional and personal contexts. In contrast, Microsoft Teams seemed to be used only by certain people in certain professions within a professional context that could exclude or impede others' participation. Furthermore, Zoom complies with the requirements set out in the GDPR[25] and holds a Cyber Essential Certificate, issued by the UK's National Cyber Security Centre.[26]

The project paid participants a £30 fee for their focus groups participation, recognising their significant work during the session. The payment made the focus groups more inclusive by not privileging those who could afford to give up two

[23] University of Oxford, 'Research Ethics (Including CUREC)' (2021) <https://researchsupport.admin.ox.ac.uk/governance/ethics#/>.

[24] Wikipedia Contributors, 'Zoom (Software)' (2021) <https://en.wikipedia.org/wiki/Zoom_(software)>.

[25] Zoom Video Communications, 'GDPR Compliance' (2020) <https://zoom.us/gdpr>.

[26] National Cyber Security Centre, 'Cyber Essentials Certificate' (2021) <https://www.ncsc.gov.uk/cyberessentials/search>.

hours of their time without financial compensation. I believe recognising participant's work during focus groups through monetary compensation is an ethical approach that appropriately values their insights and input into the research findings. The Economic and Social Research Council, as the body funding my research, provided the payments to participants. I planned to include 40 participants in the focus groups, to collect a wide range of views whilst remaining a manageable number in the context of a DPhil thesis. Several late cancellations and non-attendees meant 36 participants fully participated in 5 online focus groups. The following section will discuss the recruitment process in further detail.

Recruitment

The recruitment process took place in May 2021. One hundred and three respondents to the survey consented to be re-contacted with information on subsequent research. A first email was sent to these individuals with information on the next stage of the research. Twenty-three respondents returned their completed consent forms, but their availability and the popularity of the sessions meant not all were selected to participate. A new advert was published on the callforparticipants.org website,[27] which included details of the focus groups, the remuneration available, and the potential dates. I shared the information through my personal social media channels, which certain family and friends re-shared. One hundred forty-three individuals registered an interest through the Call for

[27] Call for Participants, 'Call for Participants' (2021) <https://www.callforparticipants.com/?re=callforparticipants.com>.

Participants webpage. Of those, eighty-six individuals returned a completed consent form and were thus able to participate.

The spaces were allocated based on individuals' availability. I ran sessions at differing times, including evenings and weekends, to best incorporate a range of people. Overall, 40 participants were recruited into five focus groups, of which 36 fully participated. The decision to host 5 focus groups consisting of 40 participants was based on research regarding the optimal number of focus group participants to facilitate rich debate and discussion best.[28]

The Set-Up

The research suggested that synchronous (real-time) focus groups should contain between three and eight participants.[29] These focus groups ranged in size from six to eight participants. The focus groups were advertised as lasting for two hours to collect as much data as possible without running for long as to risk participants dis-engaging, impacting the data quality.

The online focus groups took place through Zoom, a free and familiar software programme deemed the most accessible option. Participants were sent instructions prior to the focus groups regarding the downloading and use of Zoom. I also offered the participants the option of a trial meeting where they could test their set-up. The video and sound of the focus groups were recorded, facilitating the later transcription of both verbal and non-verbal forms of communication and displays of emotion. Though the preference for participants to use their cameras

[28] Richard A Krueger and Mary Anne Casey, *Focus Groups: A Practical Guide for Applied Research* (Sage 2015).

[29] Fielding (n 8).

was emphasised, participants could still participate without their cameras if that was their preference. Whilst it meant that I could not collect data on participants' physical responses, their participation would still enable the collection of valuable insights.

Participants were sent the survey from my quantitative research to complete before their focus group to ensure all had the same, minimal background knowledge to participate. The survey collected demographic data from participants. This quantitative aspect of the focus groups was not intended to provide the same level of robust evidence as the main quantitative research. Instead, it could provide insights into whether differences in the views expressed by survey participants and focus group participants aligned with any differences in the demographics. Furthermore, the demographic data collected could help determine whether participants with certain characteristics were more prone to articulating similar views. The participants were predominantly under thirty (55.5%,) as demonstrated in Table A.5.

Table 5: Age of Focus Group Participants

	Frequency	%
18 - 24	8	22.2%
25 - 30	12	33.3%
31 - 40	9	25.0%
41 - 50	3	8.3%
51 - 60	3	8.3%
61 - 70	1	2.8%

Table A.6 demonstrates that there was an even split between those self-identifying as male and female.

Table 6: Gender of Focus Group Participants

	Frequency	%
Male	18	50.0%
Female	18	50.0%

Of those providing their ethnicity, 44.4% identified as White British, 27.8% identified as Asian, and 13.9% identified as White Other. Other ethnicities selected by participants were African, Arab, Other and 1 participant declined to answer this question, as demonstrated in Table A.7.

Table 7: Ethnicity of Focus Group Participants

	N	%
White British	16	44.4%
White Other	5	13.9%
Asian	10	27.8%
African	2	5.6%
Arab	1	2.8%
Other	1	2.8%
Prefer not to say	1	2.8%

The survey findings drove the decision to collect participants' ethnicity. The survey data suggested that those from a minority background were more likely to

find the collection of ethnicity data for research as inappropriate.[30] Furthermore, previous research in this area has suggested that those from a minority background are more likely to be concerned about the potential for bias in such uses of data. [31] This finding is reflected in this research. Most participants (52.8%) were in full-time work, with the rest of the participants comprising students, those in part-time work, the unemployed, and the retired. As one of the provocations concerned reusing administrative data in research designed to explore inequalities in the job market and support individuals in their job search, I thought participants' employment status might be of interest when interpreting the findings. The provocations and their role in the questioning route are discussed in more detail in the next section.

The Questioning Route

The focus groups' questioning route was developed in response to the survey findings and the legislative approaches analysed in Part I. My interest was in exploring in richer detail, through participants own words and experiences, some of the divergences suggested by the research. This included divergences towards the identifiability of data and the importance of the actors in determining appropriate reuses of data in research. I also sought to explore the relationship between different factors in greater detail, as a point raised in the survey's free-

[30] Jessica Stockdale, Jackie Cassell and Elizabeth Ford, "Giving Something Back': A Systematic Review and Ethical Enquiry into Public Views on the Use of Patient Data for Research in the United Kingdom and the Republic of Ireland' (2019) 3 Wellcome Open Research 6.

[31] Daniel Cameron, Sarah Pope and Michael Clemence, 'Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes' (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/>.

text responses. Finally, I wanted to explore participants' construction of key definitions within the law and the views, experiences, and beliefs justifying their approaches. The aim was to respond to the thesis' second research question regarding the identification of divergences and their potential causes. The complete questioning route is provided in Annex C.

After a consideration of the different ways to prompt discussion within focus groups that would best respond to the research questions, I decided to use both provocations and open-ended questions within the focus groups. My work as a Policy Lab intern inspired the use of 'provocations', previously used to investigate citizens' views towards open justice.[32] The intention of the 'provocations' was to allow participants to think about and discuss the different aspects of this data processing practice, such as the actors, purpose, identifiability of the information in a more contextually-sensitive way. The survey only allowed participants to respond to each factor in isolation. However, the free-text responses highlighted the importance of the combination of factors in determining the appropriateness of a data processing activity. The use of provocations provided participants with an accessible 'entry point' to some of the complex issues at the heart of this research. They enabled participants to discuss the different aspects of the provocations that may influence their view, allowing for a more fine-grained approach than possible in the survey. This group interaction helped to unpack and interrogate the influence of these different factors, allowing for greater subtlety in the participants' responses. The provocations provided a means for participants'

[32] Rachel Bruce, 'Using Speculative Design to Explore the Future of Open Justice - Policy Lab' *openpolicy.blog.gov.uk* (November 1, 2019)

responses where they have little knowledge of direct experiences of the topic discussed during the focus groups. Furthermore, the use of provocations provided an opportunity for participants to discuss issues from a non-personal perspective, which may have encouraged them to contribute without feeling pressured to share personal information.

The provocations needed to appear possible and not outlandish to participants and incorporate the potential divergences identified through the research. The provocations drew from existing projects published on the Administrative Data Research's website and were drafted to enable participants to discuss the necessity of control and consent, the purpose of data processing, and the actors involved. The provocations were written to be as accessible, by minimising technical terminology and basing them around familiar topics such as employment and the environment. No provocation can offer a workable representation of all the relevant factors, but the resulting insights can respond to the issues raised by the research questions. The provocations provide a 'snapshot' of a situation that gives participants the space to interpret the context.[33] They cannot fully reflect participants' real-life experiences, especially in this a context where participants are unlikely to have an active role in reuses of administrative data for research. Nevertheless, it allows participants to contribute and develop their ideas. Provocations, as a tool, help shed light on individuals' perceptions, beliefs, and attitudes and help clarify principles employed, rather than act as a direct reflection of what happens. Their use uncovers some of the values, beliefs,

[33] Rhidian Hughes, 'Considering the Vignette Technique and Its Application to a Study of Drug Injecting and HIV Risk and Safer Behaviour' (1998) 20 Sociology of Health & Illness 381.

norms, and definitions by which participants may deem the actions of those involved in reusing data in research as appropriate and desirable. In doing so, the provocations would help explore participants' perception towards the legitimacy of the law in this context, understood as that assumption that the actions are desirable based on their socially constructed norms, values, beliefs, and definitions.[34] Therefore, the provocations helped elucidate some of the nebulous concepts at the heart of this thesis to explore better the divergences between the legal governance and societal expectations.

Nevertheless, the focus groups provided a unique opportunity within the empirical work to have participants respond with more freedom and build upon what others were saying. Therefore, open-ended questions explored some key terms within the GDPR and DEA legislative processes that helped justify a permissive approach towards the governance of reusing administrative data in research. The open questions could elucidate what participants thought and shed light on the reasoning and experiences they drew upon to support their views. It was also important to create a space where participants could potentially provide data on points that the research had not raised thus far.

The Focus Groups

For the focus groups, the Centre for Socio-Legal Studies, of which I am a member, kindly allowed me to use their business Zoom account for the focus group sessions. To best protect the participants' privacy and to keep the data secure, the recordings were saved to my personal laptop, rather than the cloud. This ensured

[34] Mark Suchman, 'Managing Legitimacy: Strategic and Institutional Approaches' (1995) 20 Academy of Management 571.

I was the only one able to access the files. This also meant that I transcribed the groups myself, rather than relying upon the transcription software offered by Zoom. Access to the focus groups was password-protected to prevent unwanted individuals from joining. The files were also password protected and kept on a password-protected laptop to ensure all information was secure.

During the focus group sessions, I requested that participants appear on screen where possible. Nevertheless, some participants had to resort to turning their cameras off to maintain a stable connection and others did not switch on their cameras at all. Overall, the focus groups did not suffer from any major technical difficulties. One participant's extremely poor connection meant there were difficulties transcribing their contributions though the vast majority was transcribed. A couple of participants lost connection to their session and had to re-join but quickly caught up with the discussions that had taken place during their brief absence. To proactively respond to the potential for technical issues, a colleague[35] joined the sessions. The colleague was introduced at the start of each session and their role was explained to participants. They remained muted with their camera off for the duration of the focus group. Participants could reach out to them where they were facing problems and re-enter the group if their connection was interrupted. Though it introduced another individual into the session, I felt it would be too difficult to facilitate the focus groups and deal with technical problems simultaneously.

[35] My sincere thanks to Imogen Holbrook for all her help during my research.

Each focus group commenced with an introduction which provided a brief recap of the information given to participants during the survey, including the definition of administrative data and the notion of 'reusing' data. Participants were reminded that my interest was in their views and, as such, no answer was 'right' or 'wrong' and nor was a consensus required. Each focus group started with participants introducing themselves to the group and what they thought of when they heard the word 'research.' Doing so would help participants become comfortable with the focus group format and provide an opportunity for participants to respond through their diverse experiences.

The participants then discussed the provocations, which were developed several times to explore how changes in different factors affected participants' views towards the appropriateness of the given data processing activity. The provocations were read to participants and provided to them in writing via the chat function of Zoom. Participants had 5 minutes to consider the provocation, ask for any clarifications, and consider their feelings towards the data use. The first provocation required participants to consider the following reuse of linked administrative data

> *A researcher at a university wanted to better understand wage and employment issues and inequalities in the job market. The researcher accessed and linked together government-held data about earnings, businesses, and census data. This new combination of data could result in individuals being identified. The research findings were used by the Government when implementing policies to better support people getting work and remaining employed.*

After presenting the provocation, I asked participants if they expected their data to be reused in this way and how they felt about administrative data being

used in this way. The provocation was then altered to capture any changes in their responses to shed light on the relevance of the different aspects and the interactions between these aspects. The changes in this provocation focussed on the outcomes of this data reuse to try and gain insights into participants' constructions of research. I asked participants how their views would change if there were no changes to government policy because of the research. The final shift introduced commercial entities to the provocations, to explore how participants felt about potential private sector involvement, even where this was limited to the findings, rather than the raw, potentially identifiable, data. This would enable participants to discuss how far such involvement could support the 'public interest' that may exist through such research, their views towards the market-making framing of the law as discussed in Part I, and how far the use of potentially non-identifiable data would render the reuse of their data appropriate.

The second provocation concerned the reuse of administrative data to support environmental policies.

> *For the next scenario, we're looking at a team of university researchers evaluating how people can be encouraged to adopt more environmentally friendly lifestyles. The project is sponsored by a charity interested in supporting people to become more eco-friendly. The researchers are given access to several government-held datasets and use group categories such as age or postcode to analyse the data. The researchers believe their findings could improve the government's environmental policies and be of interest to energy companies.*

The provocation explored the impact of sharing non-identifiable information and participants' views towards the involvement of charities. The provocation was developed so that the data unexpectedly became identifiable through the linkage and analysis undertaken by the researcher. A further development concerned the

outcome of such research and explored participants' constructions of the 'public interest' arising from such research where it led to financial savings but no change to the environmental impact. In doing so, it sought to explore how far the economic interests identified in Part I as a justification for the legislative approaches were also present in the focus group discussion. The provocations could enable insights into what participants deemed appropriate and desirable in this context and what norms, values, beliefs, and definitions underpinned their responses. This would respond to the second question by enabling divergences between participants responses and the approaches of the DEA and GDPR to be identified and explored.

The focus groups concluded with a series of open-ended questions,[36] which drilled down into some of the potential divergences identified in the quantitative research to explore why such divergences exist. They facilitated discussion on the role of consent and control over one's own data, participants' constructions of research, and how they felt about the law potentially allowing things that they felt to be inappropriate. The purpose of the open-ended questions was the same as the provocations, providing ways for participants to discuss their views and justify and develop their opinions, shedding light on divergences between their views and the law and the underlying reasons for these divergences. Unlike the provocations, open-ended questions enable participants to draw upon a wider range of their personal experiences and introduce new topics that may provide new insights.

[36] The full questioning route is available in Annexe C

CONCLUSION

This Annex has presented in further detail the justifications for the empirical approach and how, practically, I administered the survey and ran the online focus groups. As a sequential mixed methods piece of research, the quantitative findings, existing research, and the freedom for contributions from participants based on their own experiences, shaped the qualitative research. The empirical findings thus provide novel insights into the divergences between the legislative responses of the GDPR and the DEA and participants' views towards the appropriateness of reusing administrative data in research. By drawing upon Bacchi's framework,[37] the analysis enabled a richer understanding of why these divergences occur by exploring competing discourses, shaping different problematisations that prompted different understandings of appropriate reuses of data in research. In doing so, it provided further insights as to the legitimacy of the law governing reuses of administrative data in research, suggesting the GDPR and DEA not only lacked legitimacy, but that legitimacy would be practically impossible to secure, due to the diversity of participants responses. This allowed for the exploration of the potential responses to these divergences found in Chapter 9.

[37] Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009).

ANNEX B: SURVEY

How should data collected about you by the UK government be used in research?

Page 1: Page 1

How should data collected by the UK government about you be used?

CUREC Approval Reference: R71584/RE001

Your Participation

Welcome to this online survey. This survey aims to learn more about your thoughts on data collected by the UK government about you (your "administrative data") being used in social science research. Social science research tries to use administrative data to find new insights about the challenges facing society today. Examples of administrative data include tax data, employment data, housing data, access to benefits, and your educational data. You have been invited to participate as you are over the age of 18 and living in the UK.

During the online survey, you will be asked for your opinion on how acceptable you would find it for your administrative data to be used in research, and whether this varies depending on the type of information being shared or the institutions using the data. The survey should take no more than 15 minutes to complete and your participation is entirely voluntary. You may stop the survey at any time without giving any explanation by exiting the survey. Please contact Hannah Smith (Hannah.Smith@law.ox.ac.uk) if you have any questions before deciding to take part.

Your responses will be used in my research, and will not be shared with any third parties, including the UK government. You do not need to provide your email to participate in this survey. There is a free draw for a voucher for those who complete the survey, but you must provide your email so you can be contacted if you win. You will be asked if you want to be contacted with further information on my later research, but this is entirely optional. You can choose whether to give your email solely to enter the draw or to be informed of my future research.

Please read through this information before agreeing to participate by ticking the 'yes' box below.

How will my data be used?

Your data is treated with the strictest confidentiality and in accordance with data protection laws. This means your IP address will not be stored and your responses and any information you give will be stored in password protected files. Your responses may be used in academic publications,

but you will not be personally identifiable. Research data will be stored for a minimum of three years after publication.

Who will have access to my data?

The University of Oxford is legally responsible for overseeing the use of your data in this research project. Information about your rights with respect to your personal data is available from https://compliance.admin.ox.ac.uk/individual-rights.

This survey is for a doctoral project. The Principal Researcher is Hannah Smith, under the supervision of Professor Jane Kaye at the Centre for Health, Law, and Emerging Technologies at the University of Oxford.

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee R71584/RE001.

Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this study, please speak to Hannah Smith (Hannah.Smith@law.ox.ac.uk) or her supervisor, Professor Jane Kaye (Jane.Kaye@law.ox.ac.uk). We will respond to you within 10 working days.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible: Chair, Social Sciences & Humanities Interdivisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

If you are over 18 and have read the information above and agree to participate with the understanding that the data (including any personal data) you submit will be processed accordingly, please check the relevant box below to get started

- ○ Yes, I agree to take part
- ○ No, I do not agree to take part

Page 2: Types of Data

In your opinion, how acceptable would it be for these different types of data, which the UK government may collect about you for administrative purposes, to be reused in research?

	Very Acceptable	Quite Acceptable	Neither Acceptable nor Unacceptable	Quite Unacceptable	Very Unacceptable	Not Applicable
Your name	○	○	○	○	○	○
Your address	○	○	○	○	○	○
Your date of birth	○	○	○	○	○	○
Your ethnicity	○	○	○	○	○	○
Your income	○	○	○	○	○	○
Your religion	○	○	○	○	○	○
Your education history	○	○	○	○	○	○
Your employment history	○	○	○	○	○	○
Your immigration history	○	○	○	○	○	○
The taxes you have paid	○	○	○	○	○	○

Information about your access to benefits provided by the UK government	○	○	○	○	○	○
Your housing situation e.g. renting or homeowner	○	○	○	○	○	○
The vehicles that you own	○	○	○	○	○	○
Your use of public transport	○	○	○	○	○	○
Your use of public libraries	○	○	○	○	○	○
Your use of public parks	○	○	○	○	○	○
Your use of Council owned gyms, leisure centres, and facilities	○	○	○	○	○	○

Page 3: How identifiable would you like to be?

The ability to identify you from your administrative data depends on the form of data that is shared by the UK government with researchers. Based on this, please give your opinion on how acceptable would it be for the UK government to give researchers access to the following forms of data?

	Very Acceptable	Quite Acceptable	Neither Acceptable nor Unacceptable	Quite Unacceptable	Very Unacceptable
Anonymous data so that I cannot be identified	○	○	○	○	○
Identifiable data so researchers can access data that identifies who I am	○	○	○	○	○
Data that strongly reduces the possibility of me being identified, but does not make it impossible	○	○	○	○	○
Whatever form of data is most useful for research	○	○	○	○	○

Page 4: Interested Organisations

How acceptable would it be for the following organisations to use administrative data, originally collected by the different departments of the UK government, for research purposes?

	Very Acceptable	Quite Acceptable	Neither Acceptable nor Unacceptable	Quite Unacceptable	Very Unacceptable
Central government e.g. HM Revenue & Customs and the Department for Work and Pensions	○	○	○	○	○
Local government e.g. County Council and City Councils	○	○	○	○	○
Academics at UK universities	○	○	○	○	○
Academics at overseas universities	○	○	○	○	○
Charities	○	○	○	○	○
Community Organisations, which aim to benefit certain groups of people in a community but are not charities	○	○	○	○	○
Businesses for their own commercial purposes	○	○	○	○	○

Business when working in partnership with the UK government					

Are there any organisations that you wouldn't want to be able to access and use your administrative data for research purposes? *Optional*

Page 5: Linking Data

How acceptable would it be for the types of data listed in the first question to be linked together and used in research? For example, a researcher may want to link together your address and employment history or your housing situation and use of public transport as part of their work.

	Very acceptable for it to be linked with other data	Quite acceptable for it to be linked with other data	Neither acceptable nor unacceptable attitude towards it being linked with other data	Quite unacceptable for it to be linked with other data	Very unacceptable for it to be linked with other data
Your name	○	○	○	○	○
Your address	○	○	○	○	○
Your date of birth	○	○	○	○	○
Your ethnicity	○	○	○	○	○
Your income	○	○	○	○	○
Your religion	○	○	○	○	○
Your education history	○	○	○	○	○
Your employment history	○	○	○	○	○
Your immigration history	○	○	○	○	○
The taxes you have paid	○	○	○	○	○
Information about your access to benefits provided by the UK government	○	○	○	○	○

Your housing situation e.g. renting or homeowner					
The vehicles that you own					
Your use of public transport					
Your use of public libraries					

Are there any types of data you specifically wouldn't want to be linked together? You can refer to other types of data not included in this survey. *Optional*

Page 6: Purposes

In your opinion, how acceptable would it be for your administrative data to be used for the following purposes?

	Very Acceptable	Quite Acceptable	Neither Acceptable nor Unacceptable	Quite Unacceptable	Very Unacceptable
To improve my life	○	○	○	○	○
To improve the lives of others	○	○	○	○	○
To help the government see where it can improve public services	○	○	○	○	○
To help the government see where it could save money by making public services more effective or efficient	○	○	○	○	○
To increase our knowledge of the challenges facing society	○	○	○	○	○
For any purpose the government would like	○	○	○	○	○

Page 7: Demographic Information

Finally, we would like to collect some demographic information about you. This section is optional and you can choose to answer as little or as much as you want.

Age

- 18 - 24
- 25 - 30
- 31 - 40
- 41 - 50
- 51 - 60
- 61 - 70
- Over 70
- Prefer not to say

Gender

- Male
- Female
- Intersex
- Non-binary
- Other
- Prefer not to say

Which ethnicity do you most identify with?

- White British
- White Other
- Asian

- African
- Caribbean
- Arab
- Mixed
- Other
- Prefer not to say

If you answered 'Other', please could you give further details

Employment Status

- Full-time work
- Part-time work
- Unemployed
- Retired
- Student
- Other
- Prefer not to say

If you answered 'Other' please could you give further details.

If you would like to be entered into the free prize draw or be contacted about my future research, please enter your email below. This is completely optional and you can choose what your email can be used for in the following questions. *Optional*

Would you be happy for your email to be used to enter you into a free draw for a gift voucher?

- ○ Yes, please
- ○ No, thank you

Would you be happy to be contacted about being involved with future research? There is no obligation to take part or respond.

- ○ Yes, please
- ○ No, thank you

Page 8: Final page

Thank you for your participation, it is much appreciated! If you have questions, please contact Hannah Smith via Hannah.Smith@law.ox.ac.uk

---

ANNEX C: FOCUS GROUP QUESTIONING ROUTE

INTRODUCTION

- Prior to the questions, I will introduce myself, the format of the session (2 hours, about discussion with no right or wrong answers, being respectful of others' views, no obligation to stay, and how to respect others privacy by not identifying people and sharing quotes verbatim with others)

- I will introduce my field of interest - My project is looking at how data originally collected by the government might be used by researchers to find out more about how society works, and how it doesn't work. Researchers may be able to link different bits of data together and do analyse the data for new information and insights. Their findings may then be used by the government to inform their policies, though, of course, there is no guarantee of this!

1) Can everyone tell me your name and tell me what comes to mind when you hear the word 'research'?

PROVOCATIONS

- For the first part of this session, I'm going to give a couple of scenarios that involve the reuse of administrative data in for research purposes, as that is what I study! I'd like to hear if you expect your data to be used in this way and how you feel about the scenario. I'll say the scenarios and also post them in the chat so you can refer to them.

1) For our first scenario, a researcher at a university wanted to find out more about changes to people's wages during their career. The researcher

accessed and linked together links together data about earnings, businesses, and census data. This combination of data could allow individuals to be identified. Their research aimed to better understand wage and employment issues and inequalities in the job market. Their findings were used by the Government when implementing policies in better support people getting work and remaining employed.

a. Would you expect your data to be used in this way?

b. How do you feel about this use of data?

c. How would your views change if the researcher did their work, but the government didn't use the findings to inform their policy?

d. How would your views changed if the research was used by the government to justify remove some of the support offered to those who were looking for work, as the research suggested it wasn't particularly effective.

2) We're now moving onto our second scenario, here a researcher was interested in the evaluating housing policies. They were given access to anonymous data, making it impossible to identify individuals. Their research discovered new information about the effectiveness of the support given to those at risk of homelessness, but the government didn't use them to inform their policies in this area.

a. Would you expect your data to be used in this way?

b. How do you feel about this use of data?

c. How would your views change if the researcher found that their analysis of the data meant some people became identifiable – they were using some cutting-edge technologies in their research.

d. How would your views change if the research was undertaken by a business that provided social housing, rather than a university?

OPEN-ENDED QUESTIONS

1) Some of you disagreed with the uses of data we've just discussed. How do you feel about the possibility of the law permitting things that you wouldn't expect or that you oppose?

2) None of these scenarios required the consent of individuals whose data were reused – how do you feel about not having a say in how data may be used in these circumstances?

    a. Would you expect your consent to be asked for every time a researcher might want to access your data?

        i. What is it that you like about this?

        ii. What do you dislike about this?

3) How do you feel about others deciding whether or not it is appropriate for your data to be reused for research purposes?

    a. What do you like/dislike about this?

    b. How far do you think they represent your best interests?

    c. Do you think they consider other interests? Like what?

4) To what extent do you think you have enough control over your data?

    a. Would you like to be more involved in deciding how your data might be used?

        i. In what ways?

    b. How likely are you to oppose your data being used for a purpose you disagree with?

5) What do you think to the types of research that we've discussed today?

    a. How far do you think it's a good thing? Something we should promote? Or do you see it as perhaps risky or unhelpful?

BIBLIOGRAPHY

Ada Lovelace Institute, "Participatory Data Stewardship" (2021) <https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/> accessed September 9, 2021

——, "The Role of Good Governance and the Rule of Law in Building Public Trust in Data-Driven Responses to Public Health Emergencies" (2021) <https://www.adalovelaceinstitute.org/project/role-good-governance-rule-law-public-trust-data-driven-responses-public-health-emergencies/> accessed September 10, 2021

ADR UK, "ADR UK Annual Report 2019–20" (2020) <https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADR_UK_annual_report_2019_20.pdf> accessed July 30, 2021

——, "Data First: Harnessing the Potential of Linked Administrative Data for the Justice System" (2020) <https://www.adruk.org/our-work/browse-all-projects/data-first-harnessing-the-potential-of-linked-administrative-data-for-the-justice-system-169/> accessed July 30, 2021

——, "Homelessness Data England: Linking Local Authority Data to Evaluate Homelessness Policy" (2020) <https://www.adruk.org/our-work/browse-all-projects/homelessness-data-england-linking-local-authority-data-to-evaluate-homelessness-policy-267/> accessed July 30, 2021

——, "How Do We Work with the Public?" (2020) <https://www.adruk.org/our-mission/working-with-the-public/> accessed August 31, 2021

——, "Informing Policies to Tackle Inequalities" (2021) <https://www.adruk.org/our-mission/our-impact/article/informing-policies-to-tackle-inequalities-390/> accessed July 23, 2021

——, "Understanding Children's Lives and Outcomes in Scotland" (January 27, 2020) <https://www.adruk.org/our-work/browse-all-projects/understanding-childrens-lives-and-outcomes-in-scotland-140/> accessed July 30, 2021

Aitken M, Cunningham-Burley S and Pagliari C, "Moving from Trust to Trustworthiness: Experiences of Public Engagement in the Scottish Health Informatics Programme" (2016) 43 Science and Public Policy 713

Aitken M, de St. Jorre J, Pagliari C, Jepson R and Cunningham-Burley C, "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 <https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-016-0153-x> accessed January 25, 2021

Aitken M, Porteous C, Creamer E, and Cunningham-Burley S, "Who Benefits and How? Public Expectations of Public Benefits from Data-Intensive Health Research" (2018) 5 Big Data & Society

Aitken M, McAteer G, Davidson S, Frostick C, and Cunningham-Burley S, "Public Preferences Regarding Data Linkage for Health Research: A Discrete Choice Experiment" (2018) 3 International Journal of Population Data Science <https://ijpds.org/article/view/429> accessed January 25, 2021

Aitken M, Toreini E, Carmichael P, Coopamootoo K, Elliott K, and van Moorsel A, "Establishing a Social Licence for Financial Technology: Reflections on the Role of the Private Sector in Pursuing Ethical Data Practices" (2020) 7 Big Data & Society 1

Archibald T, "What's the Problem Represented to Be? Problem Definition Critique as a Tool for Evaluative Thinking" (2019) 41 American Journal of Evaluation 6

Article 29 Data Protection Working Party, "Opinion 4/2007 on the Concept of Personal Data" (2007)

Arvind TT and McMahon AM, "Responsiveness and the Role of Rights in Medical Law: Lessons from Montgomery" (2020) 28 Medical Law Review 445 <https://ezproxy-prd.bodleian.ox.ac.uk:2527/medlaw/article/28/3/445/5809184> accessed June 10, 2021

Ausloos J and Dewitte P, "Shattering One-Way Mirrors – Data Subject Access Rights in Practice" (2018) 8 International Data Privacy Law 4

Bacchi C and Goodwin S, *Poststructural Policy Analysis a Guide to Practice* (Palgrave Macmillan 2016)

Bacchi C, *Analysing Policy: What's the Problem Represented to Be?* (Pearson 2009)

——, *Women, Policy and Politics: The Construction of Policy Problems* (Sage 2011)

——, "Why Study Problematizations? Making Politics Visible" (2012) 02 Open Journal of Political Science 1

——, *Poststructural Policy Analysis: A Guide to Practice* (Palgrave Pivot 2016)

——, "'Are "Problem Representations" (WPR) Problematizations?'" (2018) <https://carolbacchi.com/2018/06/11/are-problem-representations-wpr-problematizations/> accessed February 21, 2021

Bache G, Flear ML and Hervey TK, "The Defining Features of the European Union's Approach to Regulating New Health Technologies'" in Mark Flear and others (eds), *European Law and New Health Technologies* (Oxford University Press 2013)

Baier M (ed), *Social and Legal Norms: Towards a Socio-Legal Understanding of Normativity* (Routledge 2016)

Ballantyne A and Schaefer GO, "Public Interest in Health Data Research: Laying out the Conceptual Groundwork" (2020) 46 Journal of Medical Ethics 610

Banakar R, "Having One's Cake and Eating It: The Paradox of Contextualisation in Socio-Legal Research" (2011) 7 International Journal of Law in Context 487

——, "Can Legal Sociology Account for the Normativity of Law?" in Matthias Baier (ed), *Social and Legal Norms: Towards a Socio-Legal Understanding of Normativity* (Routledge 2016)

Baker C and others, "COVID-19 and the Digital Divide" (*POST* December 17, 2020) <https://post.parliament.uk/covid-19-and-the-digital-divide/> accessed October 20, 2021

Barocas S and Nissenbaum H, "Big Data's End Run around Anonymity and Consent" in Julia Lane and others (eds), *Privacy, Big Data and the Public Good: Frameworks for Engagement* (Cambridge University Press 2014)

BBC News, "Edward Snowden: Leaks That Exposed US Spy Programme" *BBC News* (January 17, 2014) <https://www.bbc.co.uk/news/world-us-canada-23123964> accessed February 17, 2021

——, "Students 'May Have Been Unfairly Deported' over English Test Cheat Claims" *BBC News* (May 24, 2019) <https://www.bbc.co.uk/news/uk-48394001> accessed September 29, 2021

Bell J and others, "Balancing Data Subjects' Rights and Public Interest Research: Examining the Interplay between UK Law, EU Human Rights Law and the GDPR" (2019) 5 European Data Protection Law Review 43

Bennett Moses L, "Recurring Dilemmas: The Law's Race to Keep up with Technological Change" [2007] SSRN Electronic Journal

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=979861#> accessed February 12, 2021

Bergemann B, "The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection" in Marit Hansen and Others (eds), *Privacy and Identity Management The Smart Revolution* (Springer 2018)

Bergman M, *Advances in Mixed Methods Research* (SAGE Publications Ltd 2008)

Big Data Senior Steering Group, "The Federal Big Data Research and Development Strategic Plan" (National Science and Technology Council 2016) <https://www.nitrd.gov/pubs/bigdatardstrategicplan.pdf>

Binns R, "Analogies and Disanalogies between Machine-Driven and Human-Driven Legal Judgement" (2021) 1 Journal of Cross-disciplinary Research in Computational Law <https://journalcrcl.org/crcl/article/view/5> accessed February 23, 2021

Birkland TA, "Focusing Events, Mobilization, and Agenda Setting" (1998) 18 Journal of Public Policy 53

Blackburn P, "Outsourced and Undermined: The COVID-19 Windfall for Private Providers" *The British Medical Association* (September 8, 2020) <https://www.bma.org.uk/news-and-opinion/outsourced-and-undermined-the-covid-19-windfall-for-private-providers> accessed October 20, 2021

Bletsas A and Beasley C, *Engaging with Carol Bacchi: Strategic Interventions and Exchanges* (University of Adelaide Press 2012)

Bodó B, "The Commodification of Trust" [2021] SSRN Electronic Journal <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3843707> accessed May 25, 2021

Boswell C and Smith K, "Rethinking Policy 'Impact': Four Models of Research-Policy Relations" (2017) 3 Palgrave Communications

British Academy and Economic & Social Research Council, "A British Academy and Economic & Social Research Council Submission to the Department of Culture, Media and Sport (DCMS) Call for Views on the General Data Protection Regulation Derogations" (2017) <https://www.thebritishacademy.ac.uk/documents/104/2017-05-10_DCMS_GDPR_Derogations_submission.pdf>

——, "British Academy and ESRC Press for Shields for Humanities and Social Science Scholarship as UK Implements the New EU General Data Protection

Regulation" (2017) <https://www.thebritishacademy.ac.uk/documents/103/2017.06_-_BritAc__ESRC_GDPR_Submission_summary_3.pdf>

Brown J, "Choosing the Right Type of Rotation in PCA and EFA" (2009) <https://hosted.jalt.org/test/PDF/Brown31.pdf>

Brownsword R, "In the Year 2061: From Law to Technological Management" (2015) 7 Law, Innovation and Technology 1

Brownsword R, Scotford E and Yeung K, *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017)

Bruce Goldstein E, "Encyclopedia of Perception" (*SAGE* July 15, 2010) <https://us.sagepub.com/en-us/nam/encyclopedia-of-perception/book229708> accessed August 3, 2021

Bruce R, "Using Speculative Design to Explore the Future of Open Justice - Policy Lab" *openpolicy.blog.gov.uk* (November 1, 2019) <https://openpolicy.blog.gov.uk/2019/11/01/using-speculative-design-to-explore-the-future-of-open-justice/> accessed August 13, 2021

Bryman A, *Mixed Methods* (Sage 2006)

Buckley BS, Murphy AW and MacFarlane AE, "Public Attitudes to the Use in Research of Personal Health Information from General Practitioners' Records: A Survey of the Irish General Public" (2010) 37 Journal of Medical Ethics 50

Burke Johnson R, Onwuegbuzie AJ and Turner LA, "Toward a Definition of Mixed Methods Research" (2007) 1 Journal of Mixed Methods Research 112

Cabinet Office, "Government Digital Strategy" (2012) <https://www.gov.uk/government/publications/government-digital-strategy> accessed February 17, 2021

——, "Better Use of Data in Government - Consultation: A Government Summary of Responses" (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535063/better_use_of_data_in_government_response_final.pdf> accessed April 28, 2021

——, "Better Use of Data in Government - Consultation Paper" (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government> accessed April 28, 2021

——, “Better Use of Data in Government Consultation Responses - Section 2” (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543828/better_use_of_data_consultation_response_section_2.pdf> accessed April 29, 2021

——, “Better Use of Data in Government Consultation Responses: Section 3” (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543829/better_use_of_data_consultation_response_section_3.pdf> accessed April 29, 2021

——, “Better Use of Data in Government Consultation Responses: Section 5” (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543832/better_use_of_data_consultation_response_section_5.pdf>

——, “Better Use of Data in Government Consultation Responses: Section 6” (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543835/better_use_of_data_consultation_response_section_6.pdf>

——, “Better Use of Data in Government Consultation Responses: Section 7” (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543836/better_use_of_data_consultation_response_section_7.pdf>

——, “Better Use of Data in Government Consultation Responses: Section 8” (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/543837/better_use_of_data_consultation_response_section_8.pdf>

——, “Better Use of Data in Government” (2016) <https://www.gov.uk/government/consultations/better-use-of-data-in-government> accessed February 20, 2021

——, “Data Sharing Proposals Paper” (2016) <https://docs.google.com/document/d/1d9aoCe9e76yMtNyfkv2QjEAQOh6dyuCYdCeWb4BXFNY/edit> accessed April 29, 2021

Call for Participants, "Call for Participants" (2021) <https://www.callforparticipants.com/?re=callforparticipants.com> accessed July 26, 2021

Camden Council, "Camden Health and Care Citizens' Assembly" (2020) <https://www.camden.gov.uk/health-and-care-citizens-assembly> accessed September 8, 2021

Cameron D, Pope S and Clemence M, "Dialogue on Data: Exploring the Public's Views on Using Administrative Data for Research Purposes" (2014) <https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/> accessed February 21, 2021

Card DE and others, "Expanding Access to Administrative Data for Research in the United States" [2010] SSRN Electronic Journal <https://dx.doi.org/10.2139/ssrn.1888586> accessed June 29, 2020

Carter P, Laurie GT and Dixon-Woods M, "The Social Licence for Research: Why Care.data Ran into Trouble" (2015) 41 Journal of Medical Ethics 404

Case C-362/14 Maximillian Schrems v Data Protection Commissioner [2015] ECR I-650

Cavanillas JM, Curry E and Wahlster W (eds), *New Horizons for a Data-Driven Economy* (Cham Springer International Publishing 2016)

Centre for Data Ethics and Innovation, "Addressing Trust in Public Sector Data Use" (2020) <https://www.gov.uk/government/publications/cdei-publishes-its-first-report-on-public-sector-data-sharing/addressing-trust-in-public-sector-data-use>

Centre For Public Impact, "Building Consensus and Compromise on Uber in Taiwan" (2019) <https://www.centreforpublicimpact.org/case-study/building-consensus-compromise-uber-taiwan> accessed August 31, 2021

Chair of the Council Working Party on Statistics to The Working Group on Information Exchange and Data Protection, "Proposals for Changes in the General Data Protection Regulation" (2012)

Cheung S, "Disambiguating the Benefits and Risks from Public Health Data in the Digital Economy" (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720933924> accessed January 28, 2021

Clark VL and Ivankova NV, *Mixed Methods Research: A Guide to the Field* (Sage 2016)

Clarke A and Margetts H, "Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform" (2014) 6 Policy & Internet 393

Cohen J, "Statistical Power Analysis and Research Results" (1973) 10 American Educational Research Journal 225

Cohen JE, "Turning Privacy inside Out" (2019) 20 Theoretical Inquiries in Law 1

Collinge A, "A 'New Deal' for City Data? – London Datastore" *London.gov.uk* (January 25, 2018) <https://data.london.gov.uk/blog/a-new-deal-for-city-data/> accessed September 24, 2021

Committee on Civil Liberties, Justice and Home Affairs, "Amendments 351-601 on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)" (2013) <https://www.europarl.europa.eu/doceo/document/LIBE-AM-504340_EN.pdf?redirect>

——, "Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))" (2013) <https://www.europarl.europa.eu/doceo/document/LIBE-PR-501927_EN.pdf?redirect>

Committee on the Internal Market and Consumer Protection, "Opinion on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)" (2013) <https://www.europarl.europa.eu/doceo/document/IMCO-AD-496497_EN.pdf?redirect>

Connelly R and others, "The Role of Administrative Data in the Big Data Revolution in Social Science Research" (2016) 59 Social Science Research 1

Consolidated version of the Treaty on European Union OJ C326/1

Cotterrell R, "Why Must Legal Ideas Be Interpreted Sociologically?" (1998) 25 Journal of Law and Society 171

Council of Europe, “New Technologies: A Challenge to Privacy Protection?” (1989) <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680684607>

Council of the EU and the European Council, “The Presidency of the Council of the EU” (2021) <https://www.consilium.europa.eu/en/council-eu/presidency-council-eu/> accessed October 11, 2021

Craig P and De Búrca G, *EU Law: Text, Cases, and Materials* (Oxford University Press 2020)

Creswell JW and Plano Clark VL, *Designing and Conducting Mixed Methods Research* (3rd edn., Sage 2018)

CrowdLaw for Congress, “VTaiwan” (*congress.crowd.law*2020) <https://congress.crowd.law/case-vtaiwan.html> accessed August 31, 2021

Damschroder LJ and others, “Patients, Privacy and Trust: Patients’ Willingness to Allow Researchers to Access Their Medical Records” (2007) 64 Social Science & Medicine 223

Data Protection Act 2018

Davidson S and McLean C, “Public Acceptability of Cross-Sectoral Data Linkage” (2012) <https://www.gov.scot/publications/public-acceptability-cross-sectoral-data-linkage/> accessed January 25, 2021

Davidson S and others, “Public Acceptability of Data Sharing between the Public, Private and Third Sectors for Research Purposes” (2013) <https://www.gov.scot/publications/public-acceptability-data-sharing-between-public-private-third-sectors-research-purposes/pages/13/> accessed January 25, 2021

Dawson P and Buchanan D, “The Way It Really Happened: Competing Narratives in the Political Process of Technological Change” (2005) 58 Human Relations 845

Deligiaouri A and Suiter J, “Evaluation of Public Consultations and Citizens’ Participation in 2015 Better Regulation Agenda of the EU and the Need for a Deliberative E-Rulemaking Initiative in the EU” (2020) 22 European Politics and Society 69

de Montjoye Y-A and others, “Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata” (2015) 347 Science 536

de Vries J and others, "Knowing Who to Trust: Exploring the Role of 'Ethical Metadata' in Mediating Risk of Harm in Collaborative Genomics Research in Africa" (2014) 15 BMC Medical Ethics 62 <https://doi.org/10.1186/147269391562>

Denscombe M, "Communities of Practice" (2008) 2 Journal of Mixed Methods Research 270

Department for Business, Energy & Industrial Strategy, "The Use of Public Engagement for Technological Innovation Literature Review and Case Studies" (2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955880/use-of-public-engagement-for-technological-innovation.pdf> accessed September 9, 2021

Department for Business Innovation & Skills, "Improving Access for Research and Policy: The Government Response to the Report of the Administrative Data Taskforce" (2013) <https://www.gov.uk/government/publications/administrative-data-taskforce-report-government-response>

Department for Culture Media & Sport, "Digital Economy Bill - Overarching Impact Assessment" (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/573828/2016-11-25_Digital_Economy_Bill_-_Overarching_IA.pdf>

——, "Digital Economy Bill Factsheet - Digital Government: Research and Statistics (Clauses 56-68)" (2016) <https://www.gov.uk/government/publications/digital-economy-bill-part-5-digital-government> accessed April 28, 2021

Department for Digital, Culture, Media & Sport, "Government Response to the Consultation on the National Data Strategy" (2021) <https://www.gov.uk/government/consultations/uk-national-data-strategy-nds-consultation/outcome/government-response-to-the-consultation-on-the-national-data-strategy> accessed September 20, 2021

Department for Education, "Complete the School Census" (*GOV.* 2019) <https://www.gov.uk/guidance/complete-the-school-census> accessed February 20, 2021

Department for Work & Pensions, "Customer Information System (CIS): The Information Held about You" (2018) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/att

achment_data/file/729893/customer-information-system-the-data-held-about-you.pdf> accessed February 20, 2021

Desouza KC and Jacob B, "Big Data in the Public Sector: Lessons for Practitioners and Scholars" (2014) 49 Administration & Society 1043

Dialer D and Richter M, "Lobbying in Europe: Professionals, Politicians, and Institutions under General Suspicion?" in Doris Dialer and Margarethe Richter (eds), *Lobbying in the European Union* (Springer 2019)

Dickenson D, "The Common Good" in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017)

Digital Economy Act 2017

Digital Economy Bill. 20 October 2016 vol 615, col 246W (October 19, 2016) <https://publications.parliament.uk/pa/cm201617/cmpublic/digitaleconomy/memo/DEB57.htm> accessed April 28, 2021

Digital Economy Bill Deb 11 October 2016, vol 615col 64.

Digital Economy Bill Deb 11 October 2016, vol615 col 64.

Digital Economy Bill Deb 11 October 2016, vol615 col 82.

Digital Economy Bill Deb 13 October 2016, vol 615 col 116W (2016) <https://publications.parliament.uk/pa/cm201617/cmpublic/digitaleconomy/memo/DEB36.pdf>

Digital Economy Bill Deb 20 October 2016, vol 615, col 246W.

Digital Economy Bill Deb 20 October 2016, vol 615, col 246W (October 20, 2016)

Digital Economy Bill Deb 25 October 2016, vol616 col 308.

DIGITALEUROPE, "Response to European Commission Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data" (2009) <https://ec.europa.eu/home-affairs/sites/homeaffairs/files/what-is-new/public-consultation/2009/pdf/contributions/registered_organisations/digital_europe_en.pdf>

Dingwall R, "The Ethical Case against Ethical Regulation in Humanities and Social Science Research" (2008) 3 Twenty-First Century Society 1

Directorate-General Justice of the European Commission, "Summary of Replies to the Public Consultation about the Future Legal Framework for Protecting Personal Data" (2010) <http://ec.europa.eu/justice/news/consulting_public/0003/summary_replies_en.pdf>

Discourse (2021) <https://www.discourse.org/> accessed August 31, 2021

Dorbeck-Jung B, "Transcending the Myth of Law's Stifling Technological Innovation: How Adaptive Drug Licensing Processes Are Maintaining Legitimate Regulatory Connections" in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (1st edn., Oxford University Press 2017)

Draper NA, "From Privacy Pragmatist to Privacy Resigned: Challenging Narratives of Rational Choice in Digital Privacy Debates" (2016) 9 Policy & Internet 232

Drew C, "Data Science Ethics in Government" (2016) 374 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20160119

Dreyfus HL and Rabinow P, *Michel Foucault: Beyond Structuralism and Hermeneutics* (Routledge 2016)

Duffy R, "Analysis: Six Reasons Why Ireland Was a Landslide Yes for Repeal" *TheJournal.ie* (May 16, 2018) <https://www.thejournal.ie/reason-ireland-yes-repeal-4034286-May2018/> accessed September 8, 2021

Dumbrell NP, Adamson D and Wheeler SA, "Is Social Licence a Response to Government and Market Failures? Evidence from the Literature" (2020) 69 Resources Policy 101827

Economic & Social Research Council, "Response to the European Commission's Proposed European Data Protection Regulation (COM (2012) 11 Final)" (2013) <https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-response-to-the-european-commission-s-proposed-european-data-protection-regulation-2013/>

——, "Big Data Network" (2018) <https://esrc.ukri.org/research/our-research/big-data-network/> accessed February 21, 2021

——, "What Is Social Science?" (2018) <https://esrc.ukri.org/about-us/what-is-social-science/> accessed February 21, 2021

Eddy M and Kanter J, "Merkel Urges Europe to Tighten Internet Safeguards" *The New York Times* (July 16, 2013) <https://www.nytimes.com/2013/07/16/world/europe/merkel-urges-europe-to-tighten-internet-safeguards.html> accessed March 9, 2021

Edelman MJ, *Constructing the Political Spectacle* (The University of Chicago Press 1988)

EDRi, "Proposed Amendments to Recital 126" (April 16, 2012) <https://protectmydata.eu/recitals/recitals-121-139/recital-126/> accessed March 10, 2021

Edwards L and Veale M, "Slave to the Algorithm? Why a Right to Explanationn Is Probably Not the Remedy You Are Looking For" (2017) 16 Duke Law & Technology Review 18

Edwards MA, "The Alignment of Law and Norms: Of Mirrors, Bulwarks, and Pressure Valves" (2014) 10 FIU Law Review

Elliott D, "Data Protection Is More than Privacy" (2019) 5 European Data Protection Law Review 13

Erdos D, "Systematically Handicapped? Social Research in the Data Protection Framework" (2011) 20 Information & Communications Technology Law 83

——, "Reconciling Humanities and Social Science Research with Data Protection" (2016) <https://slideshare.net/DavidErdos1/reconciling-humanities-and-social-science-research-with-data-protection> accessed March 10, 2021

EU Charter of Fundamental Rights: Charter of Fundamental Rights of the European Union, OJ 2010 C 83/389.

EU Directive 95/46: 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.

EUROACTIV.com, "Commission Strife Risks Delaying Data Protection Overhaul" *www.euractiv.com* (January 18, 2012) <https://www.euractiv.com/section/justice-home-affairs/news/commission-strife-risks-delaying-data-protection-overhaul> accessed March 9, 2021

Europa Publications Office, "Interinstitutional Style Guide: 2.2 Preamble (Citations and Recitals)" (2019) <https://publications.europa.eu/code/en/en-120200.htm> accessed September 22, 2021

Europe-v-Facebook.org (2013) <http://europe-v-facebook.org> accessed March 9, 2021

European Commission, "Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) (Impact Assessment)" (2012) <https://www.europarl.europa.eu/cmsdata/59702/att_20130508ATT65856-1873079025799224642.pdf>

——, "Safeguarding Privacy in a Connected World a European Data Protection Framework for the 21st Century (Communication)" (COM(2012) 9 final)

——, "European Commission and Data Industry Launch €2.5 Billion Partnership to Master Big Data (Press Release)" (2014) <https://ec.europa.eu/commission/presscorner/detail/en/IP_14_1129>

European Council and Council of the European Union, "The Presidency of the Council of the EU" (2019) <https://www.consilium.europa.eu/en/council-eu/presidency-council-eu> accessed March 9, 2021

European Council, "Conclusions of the Presidency" (2000) <https://www.europarl.europa.eu/summits/lis1_en.htm#:~:text=PRESIDENCY%20CONCLUSIONS&text=The%20European%20Council%20held%20a,of%20a%20knowledge%2Dbased%20economy.> accessed March 9, 2021

European Court of Human Rights, "European Court of Human Rights" (*Coe.int* 2021) <https://www.echr.coe.int/Pages/home.aspx?p=home> accessed October 11, 2021

European Data Protection Supervisor, "A Comprehensive Approach on Personal Data Protection in the European Union (Opinion)" (2011) <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:181:0001:0023:EN:PDF>

——, "Opinion of the European Data Protection Supervisor on the Data Protection Reform Package" (2012) <https://edps.europa.eu/sites/default/files/publication/12-03-07_edps_reform_package_en.pdf>

——, "A Preliminary Opinion on Data Protection and Scientific Research" (2020) <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf> accessed September 27, 2021

European Parliament, "European Parliament Legislative Resolution of 12 March 2014 on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))" (2014) <https://www.europarl.europa.eu/doceo/document/TA-7-2014-0212_EN.html?redirect> accessed March 10, 2021

Explanatory Notes to the Digital Economy Act 2017 (2017) <https://www.legislation.gov.uk/ukpga/2017/30/notes> accessed April 28, 2021

Explanatory Notes to the Digital Economy Bill as Introduced in the House of Commons on 5 July 2016 (2016) <https://publications.parliament.uk/pa/bills/cbill/2016-2017/0045/en/17045en.pdf> accessed August 16, 2021

External Advisory Group, 'Report from the External Advisory Group for the Better Use of Data in Government Consultation' (2016) <https://web.archive.org/web/20180901171540/http://datasharing.org.uk/2016/07/15/read-the-external-advisory-group-report/index.html> accessed October 16, 2021

Faulkner A and Poort L, "Stretching and Challenging the Boundaries of Law: Varieties of Knowledge in Biotechnologies Regulation" (2017) 55 Minerva 209

Faulkner A, "Law's Performativities: Shaping the Emergence of Regenerative Medicine through European Union Legislation" (2012) 42 Social Studies of Science 753

Feilzer MY, "Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm" (2009) 4 Journal of Mixed Methods Research 6

Feliu L and Serra F, "The European Union as a 'Normative Power' and the Normative Voice of the European Parliament" in Stelios Stavridis and Daniela Irrera (eds), *The European Parliament and its International Relations* (Routledge 2015)

Field AP, *Discovering Statistics Using SPSS (and Sex and Drugs and Rock "N" Roll)* (Sage 2012)

Fielding N, *The SAGE Handbook of Online Research Methods* (Sage 2017)

Finnie R, Dubois M and Miyairi M, "How Student Pathways Affect Labour Market Outcomes: Evidence from Tax-Linked Administrative Data" (2017) <https://www.oncat.ca/sites/default/files/research-project-attachments/2016-08-final-report-university-of-ottawa-how-student-pathways-affect-labour-market-outcomes.pdf> accessed February 21, 2021

Fischer F and Forester J, *The Argumentative Turn in Policy Analysis and Planning* (Duke University Press 1993)

Fischer-Hubner S, "Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work?," *10th IFIP International Conference on Trust Management* (2016)

Flear ML and Vakulenko A, "A Human Rights Perspective on Citizen Participation in the EU's Governance of New Technologies" (2010) 10 Human Rights Law Review 661

Flear M, "The Open Method of Coordination on Health Care after the Lisbon Strategy II: Towards a Neoliberal Framing" (2009) 13 European Integration Online Papers 1

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

Fluck N, "Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data - Citizen" (https://eceuropaeu/home-affairs/sites/homeaffairs/files/what-is-new/public-consultation/2009/pdf/contributions/citizens/fluck_nick_enpdf 2009)

Fontanella-Khan J, "Brussels: Astroturfing Takes Root" *Financial Times* (June 26, 2013) <https://www.ft.com/content/74271926-dd9f-11e2-a756-00144feab7de> accessed March 9, 2021

Fougère M, Segercrantz B and Seeck H, "A Critical Reading of the European Union's Social Innovation Policy Discourse: (Re)Legitimizing Neoliberalism" (2017) 24 Organization 819

Frahm N, Doezema T and Pfotenhauer S, "Fixing Technology with Society: The Coproduction of Democratic Deficits and Responsible Innovation at the OECD and the European Commission" [2021] Science, Technology, & Human Values <https://journals.sagepub.com/doi/full/10.1177/0162243921999100> accessed September 24, 2021

Fuller LL and Winston KI, *The Principles of Social Order: Selected Essays of Lon L. Fuller* (Hart Pub 2001)

Fuller LL, *The Morality of Law*. (Yale University Press 1969)

Full Fabric, "The Impact of the GDPR in Academic Research - a Talk by Dr. David Erdos" (2017) <https://blog.fullfabric.com/the-impact-of-the-gdpr-in-academic-research-a-talk-by-dr.-david-erdos> accessed March 10, 2021

Galligan DJ, *Law in Modern Society* (Oxford University Press, 2010)

Gavaghan C, "Reproductive Technologies and the Search for Regulatory Legitimacy" in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017)

General Secretariat of the Council, "Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - Comments on Articles 83a, 83b and 83c" (2013) <https://data.consilium.europa.eu/doc/document/ST-14210-2013-REV-4/en/pdf>

——, "Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - Member States Comments on Public Sector and Chapter IX" (2014) <https://data.consilium.europa.eu/doc/document/ST-14098-2014-REV-1/en/pdf>

Godlee F, "What Can We Salvage from Care.data?" (2016) 354 BMJ 3907

Golder B, *Foucault and the Politics of Rights* (Stanford University Press, 2015)

Government Digital Service, "Government Transformation Strategy" (*GOV.UK* February 9, 2017) <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy> accessed February 17, 2021

——, "Data Ethics Framework" (2020) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/923108/Data_Ethics_Framework_2020.pdf> accessed February 21, 2021

Grecksch K and Landström C, "Drought and Water Scarcity Management Policy in England and Wales—Current Failings and the Potential of Civic Innovation" (2021) 9 Frontiers in Environmental Science

Gutwirth S, Leenes R and De Hert P (eds), *Data Protection on the Move - Current Developments in ICT and Privacy/Data Protection* (Springer 2016)

Haddow G and others, "'Nothing Is Really Safe': A Focus Group Study on the Processes of Anonymizing and Sharing of Health Data for Research Purposes" (2010) 17 Journal of Evaluation in Clinical Practice 1140

Hannah-Moffat K, "Algorithmic Risk Governance: Big Data Analytics, Race and Information Activism in Criminal Justice Debates" (2018) 23 Theoretical Criminology 453

Harris C and Chadwick L, "Edward Snowden Talks Data Exploitation at Web Summit" *euronews* (November 5, 2019) <https://www.euronews.com/2019/11/03/web-summit-2019-whistleblowers-edward-snowden-and-brittany-kaiser-set-to-speak-at-lisbon-c> accessed October 17, 2021

Harron K and others, "Challenges in Administrative Data Linkage for Research" (2017) 4 Big Data & Society <https://journals.sagepub.com/doi/10.1177/2053951717745678> accessed July 28, 2021

Hausemer P, "Participation and Political Competition in Committee Report Allocation" (2006) 7 European Union Politics 505

Hays R and Daker-White G, "The Care.data Consensus? A Qualitative Analysis of Opinions Expressed on Twitter" (2015) 15 BMC Public Health 838

HC Deb 11 October 2016, vol 615 col 61.

HC Deb 13 September 2016, vol 614 col 771.

HC Deb 13 September 2016, vol 614 col 778.

Healthcare Improvement Scotland, "Citizens' Jury: Process and Evaluation" (2019) <https://www.hisengage.scot/informing-policy/citizens-jury/process-and-evaluation/> accessed July 26, 2021

Heller B and Van Hoboken J, "Freedom of Expression: A Comparative Summary of United States and European Law" (2019) <https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/Freedom_of_Expression_TWG_Heller_May_2019.pdf> accessed March 10, 2021

Hesse-Biber S, "Qualitative Approaches to Mixed Methods Practice" (2010) 16 Qualitative Inquiry 455

Hijmans H, *The European Union as Guardian of Internet Privacy: The Story of Art 16 TFEU* (Springer 2016)

Hill EM and others, "'Let's Get the Best Quality Research We Can': Public Awareness and Acceptance of Consent to Use Existing Data in Health Research: A Systematic Review and Qualitative Study" (2013) 13 BMC Medical Research Methodology

HM Revenue & Customs, "Open Data Strategy" (2012) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/89258/open-data-strat.pdf> accessed February 21, 2021

Home Office, "Guidance – Police National Computer (PNC)" (2014) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/873384/PNC_v5.0_EXT_clean.pdf> accessed February 20, 2021

House of Lords Constitution Committee, "The Legislative Process: Preparing Legislation for Parliament Oral and Written Evidence" (2016) <https://www.parliament.uk/globalassets/documents/lords-committees/constitution/legislative-process-inquiry/evidence-volume-legislative-process-stage-1-word.pdf>

Howlett M, "Looking at the 'Field' through a Zoom Lens: Methodological Reflections on Conducting Online Research during a Global Pandemic" [2021] Qualitative Research <https://journals.sagepub.com/doi/epub/10.1177/1468794120985691> accessed March 11, 2021

Hsiao YT and others, "VTaiwan: An Empirical Study of Open Consultation Process in Taiwan"

Höchtl J, Parycek P and Schöllhammer R, "Big Data in the Policy Cycle: Policy Decision Making in the Digital Era" (2015) 26 Journal of Organizational Computing and Electronic Commerce 147

Hughes R, "Considering the Vignette Technique and Its Application to a Study of Drug Injecting and HIV Risk and Safer Behaviour" (1998) 20 Sociology of Health & Illness 381

Hull G, "Successful Failure: What Foucault Can Teach Us about Privacy Self-Management in a World of Facebook and Big Data" (2015) 17 Ethics and Information Technology 89

IBM, "SPSS Software" (*Ibm.com,* 2019) <https://www.ibm.com/analytics/spss-statistics-software>

Imagining2050, "Deliberative Futures Toolkit: Toward 'Future-Oriented' Communities and Decision-Making" (2021) <https://www.ucc.ie/en/media/projectsandcentres/imagining2050/Imagining2050 Toolkit.pdf> accessed March 30, 2021

Information Commissioner's Office, "Data Protection Rights What the Public Want and What the Public Want from Data Protection Authorities" (2015) <https://ico.org.uk/media/about-the-ico/documents/1431717/data-protection-rights-what-the-public-want-and-what-the-public-want-from-data-protection-authorities.pdf>

——, "Big Data, Artificial Intelligence, Machine Learning and Data Protection" (2018) <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

——, "Project ExplAIn Interim Report" (2019) <https://ico.org.uk/about-the-ico/research-and-reports/project-explain-interim-report/>

——, "About the ICO" (2020) <https://ico.org.uk/about-the-ico/> accessed September 10, 2021

——, "Guide to the UK General Data Protection Regulation (UK GDPR)" (2021) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/> accessed September 10, 2021

Institute for Government and others, "Digital Government during the Coronavirus Crisis" (2020) <https://www.instituteforgovernment.org.uk/publications/digital-government-coronavirus> accessed September 20, 2021

Involve, "Planning the Open Policy Process" (2014) <https://web.archive.org/web/20160814104446/http://datasharing.org.uk/2014/04/03/planning-the-open-policy-process-31-mar-2014-meeting-note/> accessed April 28, 2021

——, "Plenary Workshop with Francis Maude - Meeting Note" (October 22, 2014) <https://web.archive.org/web/20160513051853/http://datasharing.org.uk/2014/11/

07/plenary-workshop-with-francis-maude-22-oct-2014-meeting-note/> accessed April 29, 2021

——, "De-Identified Data for Research | 15 May 2014 | Meeting Note' (2014) <https://web.archive.org/web/20181129073618/http://datasharing.org.uk/category/meeting-notes/page/2/index.html> accessed April 28, 2021.

——, "De-Identified Data for Research | 13 June 2014 | Meeting Note - Data Sharing" (2014) <https://web.archive.org/web/20161109012339/http://datasharing.org.uk/2014/06/18/de-identified-data-for-research-13-june-2014-meeting-note/> accessed April 28, 2021

——, "Data for Public Benefit: Balancing the Risks and Benefits of Data Sharing" (2018) <https://www.involve.org.uk/resources/publications/project-reports/data-public-benefit-balancing-risks-and-benefits-data-sharing> accessed September 2, 2021

Ipsos MORI, "Public Attitudes to the Use and Sharing of Their Data" (2016) <https://www.slideshare.net/IpsosMORI/public-attitudes-to-the-use-and-sharing-of-their-data> accessed September 16, 2021

Irwin A, Jensen TE and Jones KE, "The Good, the Bad and the Perfect: Criticizing Engagement Practice" (2012) 43 Social Studies of Science 118

Irwin A, "STS Perspectives on Scientific Governance" in Edward J Hackett (ed), *The handbook of science and technology studies* (3rd edn., The MIT Press 2008)

Jasmontaite L and others, "Canvas White Paper 2 – Cybersecurity and Law" (2017) <https://dx.doi.org/10.2139/ssrn.3091939>

Jayanetti C, "50,000 More Homes Hit by Britain's Bedroom Tax, as Scientists Warn of COVID-19 Risk" *openDemocracy* (February 23, 2021) <https://www.opendemocracy.net/en/opendemocracyuk/50000-more-homes-hit-bedroom-tax-scientists-warn-covid-19-risk/> accessed September 16, 2021

JISC, "Online Surveys" (2021) <https://www.onlinesurveys.ac.uk/> accessed July 26, 2021

Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* [2014]

Jolliffe IT and Cadima J, "Principal Component Analysis: A Review and Recent Developments" (2016) 374 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences

<https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202> accessed July 26, 2021

Jutte DP, Roos LL and Brownell MD, "Administrative Record Linkage as a Tool for Public Health Research" (2011) 32 Annual Review of Public Health 91

Kalyanpur N and Newman AL, "The MNC-Coalition Paradox: Issue Salience, Foreign Firms and the General Data Protection Regulation" (2019) 57 JCMS: Journal of Common Market Studies 448

Kanellopoulou N, "European Commission Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data EnCoRe Project Response" (2009) <https://www.ndph.ox.ac.uk/files/research/copy_of_eudp-consulation-encore-web-1.pdf> accessed March 9, 2021

Karp P, "Government Knew about Discrepancies in Data-Matching System before Reducing Human Oversight" *the Guardian* (January 11, 2017) <https://www.theguardian.com/australia-news/2017/jan/12/centrelink-human-oversight-slashed-after-cost-benefit-analysis> accessed September 29, 2021

Kerkin S, "Designing for Legitimacy: A Systems Perspective" (2017) 15 NZJPIL 67

Kirka D, "UK Lobbying Scandal Snares Ex-PM Cameron" *AP NEWS* (April 25, 2021) <https://apnews.com/article/lobbying-david-cameron-covid-19-pandemic-europe-coronavirus-pandemic-2f84245d4901474b07fa76c8c1053b3b> accessed August 26, 2021

Kitchin R, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences* (SAGE Publications, 2017)

Kitzinger J, "Qualitative Research: Introducing Focus Groups" (1995) 311 BMJ 299

Kroes N, "Cloud Computing and Data Protection Reform" (2012) <https://ec.europa.eu/archives/commission_2010-2014/kroes/en/blog/cloud-data-protection.html> accessed March 9, 2021

——, "The Big Data Revolution" (EIT Foundation Annual Innovation Forum2013)

——, "The Economic and Social Benefits of Big Data" (Webcast Conference on Big Data 2013) <https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_13_450>

Krueger RA and Casey MA, *Focus Groups: A Practical Guide for Applied Research* (Sage 2015)

Krygier M, *Philip Selznick: Ideals in the World* (Stanford Law Books 2012)

Laerd Statistics, "Cronbach's Alpha in SPSS Statistics" (*Laerd.com*2018) <https://statistics.laerd.com/spss-tutorials/cronbachs-alpha-using-spss-statistics.php> accessed July 26, 2021

Lazaro C and Le Métayer D, "Control over Personal Data: True Remedy or Fairytale?" (2015) 12 SCRIPTed 3

Lee M, "The Legal Institutionalization of Public Participation in the EU Governance of Technology" in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press 2017)

Lemke T, "'The Birth of Bio-Politics': Michel Foucault's Lecture at the Collège de France on Neo-Liberal Governmentality" (2001) 30 Economy and Society 190

Leonard PG, "Social Licence and Digital Trust in Data-Driven Applications and AI: A Problem Statement and Possible Solutions" [2018] SSRN Electronic Journal <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261228>

Liamputtong P, *Focus Group Methodology: Principles and Practice* (Sage 2012)

Lindgren K-O and Persson T, "Input and Output Legitimacy: Synergy or Trade-Off? Empirical Evidence from an EU Survey" (2010) 17 Journal of European Public Policy 449

Lipset SM, *Political Man: The Social Bases of Politics* (The Johns Hopkins University Press 1994)

Longo J and Dobell AR, "The Limits of Policy Analytics: Early Examples and the Emerging Boundary of Possibilities" (2018) 6 Politics and Governance 5

Mageit S, "UK Government Halts GP Data Sharing Scheme after 'Mistakes' Were Made" *Healthcare IT News* (July 22, 2021) <https://www.healthcareitnews.com/news/emea/uk-government-halts-gp-data-sharing-scheme-after-mistakes-were-made> accessed August 12, 2021

Malomo F and Sena V, "Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector" (2016) 9 Policy & Internet 7

Mampuys R and Poort LM, "Controversy First: Factors Limiting the Success of Directive (EU) 2015/412 for National Decision-Making on the Cultivation of GM Crops" (2019) 11 Law, Innovation and Technology 175

Mantelero A, "The Future of Consumer Data Protection in the E.U. Re-Thinking the 'Notice and Consent' Paradigm in the New Era of Predictive Analytics" (2014) 30 Computer Law & Security Review 643

——, "Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection" (2016) 32 Computer Law & Security Review 238

Marx G and Jackson D, "Data Mining Program Designed to Predict Child Abuse Proves Unreliable, DCFS Says" *chicagotribune.com* (December 6, 2017) <https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html> accessed September 29, 2021

Mayer-Schönberger V and Padova Y, "Regime Change? Enabling Big Data through Europe's New Data Protection Regulation" (2016) 17 Science and Technology Law Review 315

McDermott Y, "Conceptualising the Right to Data Protection in an Era of Big Data" (2017) 4 Big Data & Society 1

McGuire AL and others, "DNA Data Sharing: Research Participants' Perspectives" (2008) 10 Genetics in Medicine 46

McKeon A, "#LockDownDebate: Lessons for Rapid Online Deliberation" (*Traverse.ltd*2020) <https://traverse.ltd/recent-work/blogs/online-deliberation-covid-19-lessons-weeks-12#weekone> accessed February 2, 2021

Methods and Formulas for Cross Tabulation and Chi-Square (*support.minitab.com*) <https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/basic-statistics/tables/cross-tabulation-and-chi-square/methods-and-formulas/methods-and-formulas/> accessed August 25, 2021

Micheli M and others, "Emerging Models of Data Governance in the Age of Datafication" (2020) 7 Big Data & Society <https://journals.sagepub.com/doi/full/10.1177/2053951720948087> accessed January 28, 2021

Milan S and Velden L van der, "The Alternative Epistemologies of Data Activism" (2016) 2 Digital Culture & Society 57

Mill JS, *On Liberty.* (Arcturus Publishing Ltd 1859)

Miller C, "Crossing Divides: How a Social Network Could Save Democracy from Deadlock" *BBC News* (October 25, 2019) <https://www.bbc.co.uk/news/technology-50127713> accessed September 7, 2021

Miller T and Boulton M, "Changing Constructions of Informed Consent: Qualitative Research and Complex Social Worlds" (2007) 65 Social Science & Medicine 2199

Ministry of Justice, "UK Response to the European Commission Consultation on the Legal Framework for the Fundamental Right to Protection of Personal Data" (2010)

Minkkinen M, "Making the Future by Using the Future: A Study on Influencing Privacy Protection Rules through Anticipatory Storylines" (2018) 21 New Media & Society 984

Mittelstadt B, "From Individual to Group Privacy in Big Data Analytics" (2017) 30 Philosophy & Technology 475

Mittelstadt B and Floridi L, "The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts" (2015) 22 Science and Engineering Ethics 303

Moore AD, *Privacy Rights: Moral and Legal Foundations.* (Pennsylvania State Univ 2011)

Morgan DL, *Focus Groups as Qualitative Research* (Sage Publications 2009)

Moulonguet M, "What Drives the European Parliament? The Case of the General Data Protection Regulation" [2016] Bruges Political Research Papers No 47 <http://aei.pitt.edu/73651/> accessed March 10, 2021

Mourby M 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 222

Muller SHA and others, "The Social Licence for Data-Intensive Health Research: Towards Co-Creation, Public Value and Trust" (2021) 22 BMC Medical Ethics 110

Mulligan DK, Koopman C and Doty N, "Privacy Is an Essentially Contested Concept: A Multi-Dimensional Analytic for Mapping Privacy" (2016) 374 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20160118 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124066/>

NatCen Social Research, "Public Attitudes to Data Linkage" (2018) <https://natcen.ac.uk/media/1595557/public-attitudes-to-data-linkage.pdf>

National Cyber Security Centre, "Cyber Essentials Certificate" (*Ncsc.gov.uk*2021) <https://www.ncsc.gov.uk/cyberessentials/search> accessed July 11, 2021

Nissenbaum H, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press 2009)

——, "Contextual Integrity up and down the Data Food Chain" (2019) 20 Theoretical Inquiries in Law 221

Nonet P and Selznick P, *Law and Society in Transition: Toward Responsive Law* (Transaction Publishers 2017)

Noveck BS, "Crowdlaw: Collective Intelligence and Lawmaking" (2018) 40 Analyse & Kritik 359

Nowotny H, Scott P and Gibbons M, *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty* (Polity Press 2010)

Nyrén O, Stenbeck M and Grönberg H, "The European Parliament Proposal for the New EU General Data Protection Regulation May Severely Restrict European Epidemiological Research" (2014) 29 European Journal of Epidemiology 227

O'Loughlin T and Bukowitz R, "A New Approach toward Social Licensing of Data Analytics in the Public Sector" (2021) 56 Australian Journal of Social Issues 198

OECD, *Governments Leading by Example with Public Sector Data* (2015) <https://www.oecd-ilibrary.org/content/component/9789264229358-14-en>

——, "OECD Recommendation on Public Integrity" (2017) <https://www.oecd.org/gov/ethics/recommendation-public-integrity/> accessed February 21, 2021

Oettinger GH, "Change and Innovation" (Cable Europe Annual Event 2014) <https://www.parlementairemonitor.nl/cgi-bin/as.cgi/0353000/c/start/file=/9353000/1/j9vvij5epmj1ey0/vjpeqiqzunzt?v=1&ctx=vh87km1jz6v2&start_tab0=40>

Oettinger G, "'A Digital Single Market: The Key to Europe's Industrial Leadership in the Digital Economy'" (ICT2015) <hiips://ec.europa.eu/commission/commissioners/2014-2019/Oettinger/announcements/speech-ict2015-digital-single-market-key-europes-industrial-leadership-digital-economy_en>

Ogus A, "Economics and the Design of Regulatory Law" in Stephen Copp (ed), *The legal foundations of free markets* (Institute of Economic Affairs 2008)

Ohm P, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" (2009) 57 UCLA Law Review 1701

Oswald M, "Share and Share Alike? An Examination of Trust, Anonymisation and Data Sharing with Particular Reference to an Exploratory Research Project Investigating Attitudes to Sharing Personal Data with the Public Sector" (2014) 11 SCRIPTed 245

O'Loughlin T and Bukowitz R, "A New Approach toward Social Licensing of Data Analytics in the Public Sector" (2021) 56 Australian Journal of Social Issues 198

Palazzo G and Scherer AG, "Corporate Legitimacy as Deliberation: A Communicative Framework" (2006) 66 Journal of Business Ethics 71

Paprica PA, de Melo MN and Schull MJ, "Social Licence and the General Public's Attitudes toward Research Based on Linked Administrative Health Data: A Qualitative Study" (2019) 7 CMAJ Open <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6375226/> accessed August 27, 2021

Participedia, "The Irish Citizens' Assembly" (2016) <https://participedia.net/case/5316> accessed July 26, 2021

——, "Citizens' Assembly on Brexit" (2017) <https://participedia.net/case/5166> accessed July 26, 2021

——, "Citizens' Jury" (2018) <https://participedia.net/method/155> accessed July 26, 2021

——, "Citizens' Assembly" (2020) <https://participedia.net/method/4258> accessed July 26, 2021

——, "Citizens' Assembly" (2020) <https://participedia.net/method/4258> accessed May 16, 2021

——, "#LockDownDebate: Rapid Online Deliberation on Contact Tracing" (2020) <https://participedia.net/case/6965> accessed February 1, 2021

——, "VTaiwan" (2021) <https://participedia.net/method/7387> accessed August 31, 2021

Pencheva I, Esteve M and Mikhaylov SJ, "Big Data and AI – a Transformational Shift for Government: So, What next for Research?" (2020) 35 Public Policy and Administration 24

Policy Lab, "About Policy Lab" (2014) <https://openpolicy.blog.gov.uk/about/> accessed July 27, 2021

Polis (2021) <https://pol.is/home> accessed August 31, 2021

Poort L, "Consensus & Controversies: An Interactive Legislative Approach to Animal Biotechnology in Denmark, Switzerland and the Netherlands" (PhD thesis, Tilburg University 2011) <https://research.tilburguniversity.edu/en/publications/consensus-amp-controversies-an-interactive-legislative-approach-t> accessed February 11, 2021

——, "The Tension between the Functions of Law: Ending Conflict versus Dynamics" in B van Klink, B van Beers and L Poort (eds), *Symbolic Legislation Theory and Developments in Biolaw* (Springer, Cham 2016)

Pottage A, "The Socio-Legal Implications of the New Biotechnologies" (2007) 3 Annual Reviews of Law and Social Science 321

Pound R, *Jurisprudence* (West Pub Co 1959)

Prior L, "Repositioning Documents in Social Research" (2008) 42 Sociology 821

Publications Office of the European Union, "The European Parliament: The Voice of Citizens in the European Union." (May 9, 2020) <https://op.europa.eu/en/publication-detail/-/publication/8a98b85d-932c-11ea-aac4-01aa75ed71a1/language-en> accessed September 23, 2021

QSR International, "Qualitative Data Analysis Software | NVivo" (*Qsrinternational.com* 2021) <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home> accessed October 20, 2021

Qualtrics, "Factor Analysis in Market Research" (*Qualtrics* October 16, 2020) <https://www.qualtrics.com/uk/experience-management/research/factor-analysis/?rid=ip&prevsite=en&newsite=uk&geo=GB&geomatch=uk> accessed July 26, 2021

Raab C and Koops B-J, "Privacy Actors, Performances, and the Future of Privacy Protection" in Serge Gutwirth, Yves Poullet and Paul Hert (eds), *Reinventing Data Protection?* (Springer 2009)

Reding V and Shatter A, "What EU Data Protection Can Do for Growth" (*www.eu2013.ie* 2013) <http://www.eu2013.ie/news/news-items/20130129shatterredingoped> accessed March 9, 2021

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

Rempel ES, Barnett J and Durrant H, "Public Engagement with UK Government Data Science: Propositions from a Literature Review of Public Engagement on New Technologies" (2018) 35 Government Information Quarterly 569

Research Excellence Framework, "Index of Revisions to the 'Panel Criteria and Working Methods' (2019/02)" (2020) <https://www.ref.ac.uk/media/1084/ref-2019_02-panel-criteria-and-working-methods.pdf> accessed February 21, 2021

Richards NM and King JH, "Big Data Ethics" (2014) 49 Wake Forest Law Review 393

Richardson JJ, *European Union: Power and Policy-Making* (Routledge 2006)

Rieder G and Simon J, "Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data" (2016) 3 Big Data & Society

Rieder G, "Tracing Big Data Imaginaries through Public Policy: The Case of the European Commission" in Ann Rudinow Sætnan, Ingrid Schneider and Nicola Green (eds), *The Politics of Big Brother: Big Data, Big Brother?* (Taylor Francis 2018)

Ripoll Servent A, "Protecting or Processing? Recasting EU Data Protection Norms" in W Schünemann and MO Baumann (eds), *Privacy, Data Protection and Cybersecurity in Europe* (Springer 2017)

Robling MR, "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

Romm NRA, "Employing Questionnaires in Terms of a Constructivist Epistemological Stance: Reconsidering Researchers' Involvement in the Unfolding of Social Life" (2013) 12 International Journal of Qualitative Methods 652

Rossi A, "How the Snowden Revelations Saved the EU General Data Protection Regulation" (2018) 53 The International Spectator 95

Royal Statistical Society, "Royal Statistical Society Research on Trust in Data and Attitudes toward Data Use / Data Sharing" (2014) <http://www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf>

Rubinstein I, "Big Data: The End of Privacy or a New Beginning?" (2013) 3 International Data Privacy Law 74

Rule JB, "Privacy and the Illusions of 'Balancing'" (2020) 6 European Data Protection Law Review 7

Russon Gilman H and Carneiro Peixoto T, "Digital Participation" in Stephen Elstub and Oliver Escobar (eds), *Handbook of Democratic Innovation and Governance* (Edward Elgar Publishing 2019)

Saxena N and others, "Understanding Canadians' Attitudes and Expectations Citizens' Dialogue on Privacy and the Use of Personal Information for Health Research in Canada Public Involvement Network" (2006) <http://cprn3.library.carleton.ca/documents/43383_en.pdf>

Schermer BW, Custers B and van der Hof S, "The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection" [2014] Ethics and Information Technology

Science and Technology Committee, "The Big Data Dilemma" (HC (2015-16) 468) <https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/46802.htm>

Selznick P, "'Law in Context' Revisited" (2003) 30 Journal of Law and Society 177

Simon J, Bass T and Boelman V, "Digital Democracy: The Tools Transforming Political Engagement" (2017) <https://www.issuelab.org/resources/33129/33129.pdf> accessed September 9, 2021

SlideShare, "SlideShare.net" (*www.slideshare.net*2020) <https://www.slideshare.net/> accessed August 31, 2021

Smith D and Smith T, "Using Linking to Sharpen Policy-Thinking: Early Findings from the Australian Census and Migrants Integrated Dataset" (2014) 47 Australian Economic Review 220

Solove D, "The Meaning and Value of Privacy Appeal for a Pluralistic Definition of the Concept of Privacy" *Open!* (April 2010) <https://onlineopen.org/the-meaning-and-value-of-privacy> accessed February 20, 2021

Solum LB, "Public Legal Reason" (2006) 92 Virginia Law Review 1472

Sorbie A, "Sharing Confidential Health Data for Research Purposes in the UK: Where Are 'Publics' in the Public Interest?" (2019) 16 Evidence & Policy: A Journal of Research, Debate and Practice 249

Spindler G and Schmechel P, "Personal Data and Encryption in the European General Data Protection Regulation" (2016) 7 Journal of Intellectual Property, Information Technology and E-Commerce Law 163

Starkbaum J and Felt U, "Negotiating the Reuse of Health-Data: Research, Big Data, and the European General Data Protection Regulation" (2019) 6 Big Data & Society 1

Steinmann M, Matei S and Collmann J, "A Theoretical Framework for Ethical Reflection in Big Data Research" in SA Matei and J Collmann (eds), *Ethical Reasoning in Big Dat : An Exploratory Analysis* (Springer Netherlands 2016)

Sterckx S and Cockbain J, "The Natural, the Informational, the Claimable? Human Body Material in US and European Patent Law" in B van Klink, B van Beers and L Poort (eds), *Symbolic Legislation Theory and Developments in Biolaw* (Springer, Cham 2016)

Sterckx S and others, "'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" (2015) 19 Medicine, Health Care and Philosophy 177

Stewart K and Williams M, "Researching Online Populations: The Use of Online Focus Groups for Social Research" (2005) 5 Qualitative Research 395

Stockdale J, Cassell J and Ford E, "'Giving Something Back': A Systematic Review and Ethical Enquiry into Public Views on the Use of Patient Data for Research in the United Kingdom and the Republic of Ireland" (2019) 3 Wellcome Open Research <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6402072/> accessed January 28, 2021

Stokes E, "Something Old, Something New, Something Borrowed: Emerging Health Technologies and the Continuing Role of Existing Regulations" in Mark L Flear and others (eds), *European Law and New Health Technologies* (Oxford University Press 2013)

Strandburg KJ, "Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context" in Julia Lane and others (eds), *Privacy, Big Data,*

and the Public Good: Frameworks for Engagement (Cambridge University Press 2014)

Street J and others, "Sharing Administrative Health Data with Private Industry: A Report on Two Citizens' Juries" (2021) 0 Health Expectations 1

Strycharz J, Ausloos J and Helberger N, "Data Protection or Data Frustration? Individual Perceptions and Attitudes towards the GDPR" (2020) 6 European Data Protection Law Review 407

Study in Denmark, "Denmark Launches World-Class Biobank" *studyindenmark.dk* (2019) <https://studyindenmark.dk/news/denmark-launches-world-class-biobank> accessed March 9, 2021

Suchman M, "Managing Legitimacy: Strategic and Institutional Approaches" (1995) 20 Academy of Management 571

Tabachnick BG, Fidell LS and Ullman JB, *Using Multivariate Statistics* (Boston Pearson 2019)

Taekema S, "Two Realist Idealists Theorizing the Rule of Law" (2019) 11 Hague Journal on the Rule of Law 301

Tam H, "Government with the People" in Henry Tam (ed), *Whose Government Is It?: The Renewal of State-Citizen Cooperation* (Bristol University Press 2020)

—— (ed), *Whose Government Is It?: The Renewal of State Citizen Cooperation.* (Bristol University Press 2020)

Tamanaha BZ, *A General Jurisprudence of Law and Society* (Oxford University Press 2010)

Tang, Audrey 唐鳳, "Uber Responds to VTaiwan's Coherent Blended Volition" *Medium* (May 23, 2016) <https://blog.pol.is/uber-responds-to-vtaiwans-coherent-blended-volition-3e9b75102b9b> accessed September 7, 2021

Taylor MJ and Whitton T, "Public Interest, Health Research and Data Protection Law: Establishing a Legitimate Trade-off between Individual Control and Research Access to Health Data" (2020) 9 Laws 6

Tene O and Polentsky J, "Privacy in the Age of Big Data a Time for Big Decisions" (2012) 64 Stanford Law Review

The Citizens' Assembly, “About the Citizens’ Assembly” (*Citizensassembly.ie* 2019) <https://www.citizensassembly.ie/en/about-the-citizens-assembly/> accessed September 8, 2021

The Conservative Party, “The Conservative Party Manifesto 2015” (2015) <https://issuu.com/conservativeparty/docs/ge_manifesto_low_res_bdecb3a47a0faf> accessed April 28, 2021

The Presidency of the Council of the European Union, “Data Protection Package - Report on Progress Achieved under the Cyprus Presidency” (2012) <https://data.consilium.europa.eu/doc/document/ST-16525-2012-INIT/en/pdf>

——, “Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)” (2012) <https://www.statewatch.org/media/documents/news/2012/jun/eu-council-revised-dp-position-11326-12.pdf> accessed March 10, 2021

——, “Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)” (2015) <https://data.consilium.europa.eu/doc/document/ST-10391-2015-INIT/en/pdf>

——, “Implementation of Risk-Based Approach in the General Data Protection Regulation” (The Presidency of the Council of the European Union 2013) <https://data.consilium.europa.eu/doc/document/ST-5702-2013-INIT/en/pdf>

——, “Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)” (2015) <https://data.consilium.europa.eu/doc/document/ST-10391-2015-INIT/en/pdf>

Thompson B, “Academic Research Perspective on the European Commission, Parliament and Council Texts of the Proposal for a General Data Protection Regulation” (2015) <https://wellcome.org/sites/default/files/research-perspective-data-protecton-regulation-proposal-wellcome-jul15.pdf>

Timms J, “Why Is Free Speech Important?” *Index on Censorship* (April 13, 2016) <https://www.indexoncensorship.org/2016/04/free-speech-important>

TNS BMR, “Open Data Dialogue: Final Report” (2013) <https://www.force11.org/presentation/open-data-dialogue-final-report> accessed September 16, 2021

Trinidad SB and others, “Genomic Research and Wide Data Sharing: Views of Prospective Participants” (2010) 12 Genetics in Medicine 486

——, “Informed Consent in Genome-Scale Research: What Do Prospective Participants Think?” (2012) 3 AJOB Primary Research 3

Trischler J, Dietrich T and Rundle-Thiele S, “Co-Design: From Expert- to User-Driven Ideas in Public Service Design” (2019) 21 Public Management Review 1595

Tully MP and others, “Commercial Use of Health Data—a Public ‘Trial’ by Citizens’ Jury” (2019) 3 Learning Health Systems

UCLA: Statistical Consulting Group, “A Practical Introduction to Factor Analysis: Exploratory Factor Analysis” (*Ucla.edu* 2021) <https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/> accessed July 26, 2021

UK Parliament, “Digital Economy Act 2017 Stages” (*Parliament.uk* 2017) <https://bills.parliament.uk/bills/1859/stages> accessed April 28, 2021

UK Research and Innovation, “Pathways to Impact” (*Nationalarchives.gov.uk* 2018) <https://webarchive.nationalarchives.gov.uk/20200203113953/https://www.ukri.org/innovation/excellence-with-impact/pathways-to-impact/> accessed February 21, 2021

UK Statistics Authority, “Data Sharing Code of Practice and Accreditation Criteria: Research” (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567074/Research_Code_and_Accreditation_Criteria__1___2_.pdf> accessed April 29, 2021

University of Oxford, “Research Ethics (Including CUREC)” (2021) <https://researchsupport.admin.ox.ac.uk/governance/ethics#/> accessed July 26, 2021

——, “Social Sciences & Humanities Interdivisional Research Ethics Committee (SSH IDREC)” (2021) <https://researchsupport.admin.ox.ac.uk/governance/ethics/committees/sshidrec> accessed July 26, 2021

van der Burg W, “The Emerging Interactionist Paradigm and the Ideals of Democracy and Rule of Law” in Bart van Klink, Britta van Beers and Lonneke

Poort (eds), *Symbolic Legislation Theory and Developments in Biolaw* (Springer 2016)

van der Burg W and Brom F, "Legislation on Ethical Issues: Towards an Interactive Paradigm" (2000) 3 Ethical Theory and Moral Practice 57

van der Sloot B, "Do Data Protection Rules Protect the Individual and Should They? An Assessment of the Proposed General Data Protection Regulation" (2014) 4 International Data Privacy Law 307

van de Waerdt PJ, "Information Asymmetries: Recognizing the Limits of the GDPR on the Data-Driven Market" (2020) 38 Computer Law & Security Review 105436

van Klink B, "Legislation, Communication, and Authority. How to Account for the Bindingness of Law?" in A Daniel Oliver-Lalana (ed), *Conceptions and Misconceptions of Legislation* (Springer International Publishing 2019)

van Ooijen C, Ubaldi B and Welby B, "A Data-Driven Public Sector: Enabling the Strategic Use of Data for Productive, Inclusive and Trustworthy Governance" [2019] https://www.oecd-ilibrary.org <https://www.oecd-ilibrary.org/docserver/09ab162c-en.pdf?expires=1591964721&id=id&accname=guest&checksum=DF891758A592EB42EC483635D8F6928B>

Veress J, "Modelling Civil Society's Transformational Dynamism and Its Potential Effects," *European Conference on Modelling and Simulation* (2017) <http://www.scs-europe.net/dlib/2017/ecms2017acceptedpapers/0106-fes_ECMS2017_0096.pdf>

Vezyridis P and Timmons S, "Understanding the Care.data Conundrum: New Information Flows for Economic Growth" (2017) 4 Big Data & Society 1

Vie Publique, "Convention Citoyenne Pour Le Climat : Une Expérience Démocratique Inédite" *Vie publique.fr* (May 18, 2021) <https://www.vie-publique.fr/eclairage/279701-convention-citoyenne-pour-le-climat-experience-democratique-inedite> accessed August 19, 2021

Vincent-Jones P, "Responsive Law and Governance in Public Services Provision: A Future for the Local Contracting State" (2003) 61 The Modern Law Review 362

vTaiwan, "VTaiwan.tw" (2017) <https://vtaiwan.tw/> accessed September 2, 2021

——, "Sharing Economy" (*Vtaiwan.tw*2017) <https://vtaiwan.tw/topic/sharing-economy/> accessed August 31, 2021

Vydra S and others, “Big Data Ethics: A Life Cycle Perspective” (2021) 14 Erasmus Law Review

Waind E, “Trust, Security and Public Interest: Striking the Balance” (2020) 5 International Journal of Population Data Science

Warren SD and Brandeis LD, “The Right to Privacy” (1890) 4 Harvard Law Review 193

Weiss JA, “Access to Influence” (1979) 22 American Behavioral Scientist 437

Weitzman ER, Kaci L and Mandl KD, “Sharing Medical Data for Health Research: The Early Personal Health Record Experience” (2010) 12 Journal of Medical Internet Research

Wellcome Trust, “Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data” (*Wellcome Library* July 2013) <https://wellcomelibrary.org/item/b20997358#?c=0&m=0&s=0&cv=0> accessed February 26, 2021

——, “Data Protection” *www.thetimes.co.uk* (January 2014) <https://www.thetimes.co.uk/article/data-protection-dj89jl8rb8w> accessed March 9, 2021

Wellcome Trust and others, “Protecting Health and Scientific Research in the Data Protection Regulation (2012/0011(COD)) Position of Non-Commercial Research Organisations and Academics -April 2014” (2014) <https://www.russellgroup.ac.uk/media/5288/dpr-ep-position-joint-statement-april-2014.pdf>

Welpton R and Wright M, “HMRC Datalab: Engaging with the External Research Community” (2011) <https://dam.data-archive.ac.uk/reports/research/hmrcdatalab.pdf>

Wesselink A and others, “Technical Knowledge, Discursive Spaces and Politics at the Science–Policy Interface” (2013) 30 Environmental Science & Policy 1

Westin Alan, *Privacy and Freedom* (Ig Publishing 2015)

Whitley EA, Kanellopoulou N and Kaye J, “Consent and Research Governance in Biobanks: Evidence from Focus Groups with Medical Researchers” (2012) 15 Public Health Genomics 232

Whitman JQ, “The Two Western Cultures of Privacy: Dignity versus Liberty” (2003) 113 Yale Law Journal 1151

Wikipedia Contributors, “Jan Philipp Albrecht” (*Wikipedia* June 21, 2021) <https://en.wikipedia.org/wiki/Jan_Philipp_Albrecht> accessed October 11, 2021

——, “Speech Disfluency” (*Wikipedia* October 7, 2019) <https://en.wikipedia.org/wiki/Speech_disfluency>

——, “Zoom (Software)” (*Wikipedia* July 10, 2021) <https://en.wikipedia.org/wiki/Zoom_(software)> accessed July 11, 2021

Willison DJ and others, “Consent for Use of Personal Information for Health Research: Do People with Potentially Stigmatizing Health Conditions and the General Public Differ in Their Opinions?” (2009) 10 BMC Medical Ethics

Willison DJ, “Patients’ Consent Preferences for Research Uses of Information in Electronic Medical Records: Interview and Survey Data” (2003) 326 BMJ 373

Woollard M, “Administrative Data: Problems and Benefits. A Perspective from the United Kingdom” in Adrian Duşa and others (eds), *Facing the Future: European Research Infrastructure for Humanities and Social Sciences* (SCIVERO 2014)

Wynne B, “Misunderstood Misunderstanding: Social Identities and Public Uptake of Science” (1992) 1 Public Understanding of Science 281

——, “Creating Public Alienation: Expert Cultures of Risk and Ethics on GMOs” (2001) 10 Science as Culture 445

Xafis V, “The Acceptability of Conducting Data Linkage Research without Obtaining Consent: Lay People’s Views and Justifications” (2015) 16 BMC Medical Ethics

Yiu C, *The Big Data Opportunity: Making Government Faster, Smarter and More Personal* (Policy Exchange 2012)

Zarsky T, “Transparent Predictions” (2013) 4 University of Illinois Law Review 1503

——, “Incompatible: The GDPR in the Age of Big Data” (2017) 47 Seton Hall Law Review

Zoom Video Communications, “GDPR Compliance” (2020) <https://zoom.us/gdpr> accessed July 11, 2021