

REGULATING ADMINISTRATIVE DATA SHARING FOR RESEARCH PURPOSES IN THE UNITED KINGDOM

Thesis submitted to the Faculty of Law at the University of Oxford
for the degree of Doctor of Philosophy
by Stergios Aidinlis
St Cross College

Hilary Term 2021
99,155 words

ABSTRACT

As the provision of more and more public services becomes digitalised, administrative data research, i.e. statistical research on datasets collected by the central and local government throughout its operation, has received increased attention from policy-makers over the last few decades. In the UK, the government has embarked on a series of policy and legal initiatives, as well as financial investments, to widen access to public-sector datasets for research use. Yet, empirical studies identify the existence of 'legal' and 'cultural' barriers to 'sharing' administrative data for research purposes, raising questions about the forces that regulate data access decisions by public-sector organisations responsible for managing the datasets. This thesis examines the regulation of research data access decisions in the UK public sector through a 'regulatory space analysis' lens by asking: what are the 'rules of the game' that filter access to public-sector administrative data for research purposes in the UK? How and by which actors have these rules been developed over the last four decades?

To address these questions, the thesis employs an in-depth, qualitative research design which involved semi-structured interviews (with 42 participants) and documentary data about the regulation of research data sharing decisions by specific organisations in the UK public sector. The thesis shows that research data access has been governed by rules which are incrementally established and revised via small-scale interactions, often between low-level technical staff, rather than momentous legal and policy decisions made by high-level officials or bodies. Through these interactions, the 'political' is artfully concealed behind the 'technical' and the 'commonplace'. Well-established political views, reflected in unwritten, conventional normative assumptions about the ownership of administrative data and the relevance of administrative data research for policy-making, effectively shape the 'rules of the game' that govern data access in practice.

TABLE OF ABBREVIATIONS **7**

TABLE OF CASES **7**

TABLE OF LEGISLATION **8**

TABLE OF TREATIES **9**

1. INTRODUCTION **10**

A. Regulating Administrative Data Sharing for Research Purposes in the UK: Theory and Practice **11**

B. Theorising a Research Data Sharing Decision-Making Space: Unwritten Norms and Informal Interactions **18**

C. Outline of Thesis Chapters **21**

2. THEORISING THE 'DECISION-MAKING SPACE' OF ADMINISTRATIVE DATA SHARING FOR RESEARCH IN THE UK: ACTORS, INSTITUTIONS AND INTERACTIONS **25**

A. Regulating 'Research Data Sharing Decisions' from a Spatial Perspective: Actors, Institutions and Interactions **26**

i. Regulating 'Research Data Sharing Decisions': Outcomes or Processes? **26**

ii. The 'Space' of Making Decisions about Research Data Sharing and its Dimensions **30**

B. Mapping the Actors Populating the Research Data Sharing Space **34**

i. High-level standard-setters **39**

ii. Data custodians **40**

iii. Researchers **41**

iv. Intermediary Bodies **42**

C. Institutionalisation in the Research Data Sharing Space **45**

D. Regulatory Interactions in the Research Data Sharing Space: From Macro to Micro **50**

3. STUDYING THE REGULATION OF RESEARCH DATA SHARING DECISIONS THROUGH AN IN-DEPTH, QUALITATIVE RESEARCH DESIGN **57**

A. A Micro-Level 'Anatomy' of Research Data Sharing Decisions **57**

B. Collecting Data about the Regulation of Research Data Sharing Decisions **61**

i. Choice of Field Setting **61**

ii. Selection of Data Collection Methods and Sampling **63**

iii. Interview Preparation and Ethics **66**

C. An Issue-Focused, Thematic Analysis of the Regulation of Research Data Sharing Decisions **68**

D. 'Validity', 'Reliability' and Limitations **72**

i. 'Validity' **72**

ii. 'Reliability' **73**

iii. Limitations **74**

4. BETWEEN 'EFFICIENCY' AND 'ACCOUNTABILITY': THE ACTORS WHO SHAPED THE HISTORICAL DEVELOPMENT OF ADMINISTRATIVE DATA SHARING FOR RESEARCH IN THE UK PUBLIC SECTOR 77

A. Informality: Ad-Hoc Bargaining to Access Data for Research **80**

i. Controlled access for researchers and data custodian reluctance **80**

ii. The first administrative data research projects: doing research in-house or commissioning academic researchers **84**

iii. The origins of 'independent research': demanding independence, settling for informality **90**

iv. Widening research data access: the view from the high level **95**

B. Formalisation: Creating Formal Frameworks to Streamline Research Data Sharing in the UK **96**

i. 'Independent' Research Data Access, 'Efficiency' and 'Accountability' **96**

ii. Formalising 'Independent' Research Data Access: Trailblazers and the Ambition for Wider Formalisation **102**

iii. The ADRN (2013-18) as an Agent of Systematic Formalisation **107**

C. Fragmentation: Unequal Access of Researchers to Administrative Data in the UK **110**

i. After the Administrative Data Research Network: fragmentation and the limits of formalisation **111**

ii. The eclipse of 'accountability' and the complications for independent researchers **116**

D. Towards studying research data sharing institutions: institutions & interactions **120**

5. THE LIMITED REACH OF 'LEGAL' INTERPRETATIONS OF ACTORS IN THE REGULATION OF RESEARCH DATA SHARING DECISIONS 123

A. 'Legal' Interpretations as Research Data Sharing Institutions **124**

B. The Legal Framework applicable to Research Data Sharing in the UK **127**

i. Powers to Share Data for Research Purposes in UK Law **129**

a. *Statutory gateways* **129**

b. *Implied powers* **130**

c. *Common law powers* **132**

d. *Prerogative powers* **134**

ii. Limits on Powers to Share Data for Research Purposes in UK Law **135**

a. *Limitations stemming from data protection law* **136**

b. *Limitations stemming from human rights law* **139**

C. Law's Limited Reach: Interview Data from the Four UK Nations **142**

i. England: the letter of the law and the spirit of the public benefit **144**

ii. Northern Ireland: the law barks, but does not bite **147**

iii. Wales: the law as a post-hoc justification **150**

iv. Scotland: it's not the law, it's the lawyers **154**

D. From Law's Limited Reach to the Salience of a Research Data Sharing Culture **159**

6. THE SALIENCE OF A 'RESEARCH DATA SHARING CULTURE' FOR THE REGULATION OF RESEARCH DATA SHARING DECISIONS 160

A. Perceptions of Administrative Data 'Ownership': Sharing Whose Data? **162**

i. Data as a 'public good': enabling research and scrutinizing the government **162**

ii. Data as the 'public good': establishing mechanisms of accountability **168**

iii. Data as 'custodian property': 'our' data and 'their' research **176**

iv. Data as 'custodian property': resisting accountability **180**

B. Policy Relevance of Administrative Data Research: 'What is in it for us'? **187**

i. 'Blue skies' research and the normative supremacy of policy-relevant research **188**

ii. Policy-relevance within data sharing frameworks: brigading research projects **197**

iii. Policy-relevance within data sharing frameworks: co-designing research projects **203**

C. Conclusion: The Salience of a 'Research Data Sharing Culture' for Regulating Decisions **208**

7. THE CENTRALITY OF MICRO-LEVEL INTERACTIONS IN SHAPING THE REGULATION OF RESEARCH DATA SHARING DECISIONS 210

A. The Multi-Faceted, Central Role of Data Analysts in the Research Data Sharing Space **211**

i. Analysts as experts: knowing the data, shaping the process **212**

ii. Analysts as mediators: helping researchers, streamlining data sharing processes **222**

iii. Analysts as decision-makers: blurring the lines between 'recommending' and 'deciding' **229**

B. Micro-Level Communication: Connections, Informal Communications and Small-Scale Interactions in the Research Data Sharing Space **235**

i. 'Knowing the right person': trust and relationships in the research data sharing space **235**

ii. Informal interactions: shaping data sharing frameworks far from the spotlight **240**

iii. Small-scale interactions: iteratively and incrementally shaping data sharing frameworks **246**

C. Conclusion: The Importance of Commonplace, Low-Level Interactions **252**

8. CONCLUSIONS AND RECOMMENDATIONS: THE PAST, PRESENT AND FUTURE REGULATION OF 'INDEPENDENT', DATA-DRIVEN RESEARCH 254

A. The Research Data Sharing Space: Concealing the Political Behind the Technical and the Commonplace **254**

B. The Future for the Regulation of Research Data Sharing: Designing Legal and Cultural Shifts **260**

C. Limitations and Main Directions for Future Research **262**

APPENDICES **265**
A. Participant recruitment letter **265**
B. Participant information sheet **266**
C. List of participants **269**
D. Interview guide **272**
E. Pilot survey **275**
BIBLIOGRAPHY **280**

TABLE OF ABBREVIATIONS

ADRN	Administrative Data Research Network
BEIS	Department for Business, Energy and Industrial Strategy
DfE	Department for Education
DWP	Department for Work and Pensions
ECHR	European Convention on Human Rights
ESRC	Economic and Social Research Council
HMCTS	Her Majesty's Courts and Tribunals Service
HMRC	Her Majesty's Revenue and Customs
ICO	Information Commissioner's Office
MoJ	Ministry of Justice
OECD	Organisation for Economic Co-Operation and Development
ONS	Office for National Statistics
UKSA	UK Statistics Authority

TABLE OF CASES

Attorney General v Great Eastern Railway Co (1880) 5 App Cas 473……128
D and J Nicol v Dundee Harbour Trustees (1914) 2 SLT 418…….……128
Attorney General v De Keyser's Royal Hotel Ltd (1920) AC…….……..131
Rederiaktiebolaget Amphitrite v The King (1921) 3 KB 500…….……..130
Robertson v Minister of Pensions (1949) 1 KB 227…….……..…..130
Crown Lands Commissioner v Page (1960) 2 QB 274…….……....130
Glasgow Corporation v Flint (1966) SLT 183…….……128
Leander v Sweden (1987) 9 EHRR 433……138
R v Secretary of State for the Home Department, ex parte Northumbria Police Authority (1989) QB 26……..…..……..131
Amann v Switzerland (2000) 30 EHRR 843……..138
R v Secretary of State for Health ex p C (2000) 1 FLR 627……..132
R (Daly) v Secretary of State for the Home Department (2001) UKHL 26……..138
R (Hooper) v Secretary of State for Work and Pensions (2005) UKHL 29……..131
Ward v Commissioner of Police of the Metropolis and another (2005) UKHL 32……..129

Shrewsbury and Atcham Borough Council v Secretary of State for Communities and Local Government (2008) EWCA Civ 148……..132
Common Services Agency v Scottish Information Commissioner (2008) UKHL 47…….133
R (New London College Ltd) v S. of State for the Home Department (2013) UKSC 51…..129
South Lanarkshire Council v Scottish Information Commissioner (2013) UKSC 55……...137
R (on the application of W, X, Y, and Z) v The Secretary of State for Health v The Secretary of State for the Home Department (2014) EWHC 1532 (Admin)……...…..128
Digital Rights Ireland Ltd v Minister for Communications (ECLI:EU:C:2014:238)……138
Christian Institute v Lord Advocate (2016) UKSC 51…….……..38, 63, 132, 137, 138, 139
Satakunnan Markkinapörssi Oy v Finland (2018) 66 EHRR 8…….137

TABLE OF LEGISLATION

Children Act 1989…….…….……126
Crime and Disorder Act 1998…….……..…..……126
Human Rights Act 1998…….…….……..…..……137
Adults With Incapacity (Scotland) Act 2000…….……125
Freedom of Information Act 2000…….……113, 133, 259
Local Government Act 2000…….…….……126
Education Act 2002…….……..…….……126
Regulation 2, 3(4) or 5 of the Health Service (Control of Patient Information) Regulations 2002…….……...……151
National Health Service Act (NHS Act) 2006…….……14, 126, 151
Statistics and Registration Services Act (SRSA) 2007……..……10, 47, 126, 140, 167
Prescribed Persons Act 2009…….……125, 126, 141
The Education (Pupil Information) (Wales) Regulations 2011…….125
Localism Act 2011…….130
Health and Social Care (HSC) Act 2012…….…….14, 126, 139
Care Act 2014…….126
Children and Young People (Scotland) Act (CYPA) 2014……138, 139
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)……..…..…..39, 107, 123, 125, 131, 134, 135, 136, 154, 155, 218, 240
Digital Economy Act (DEA) 2017…….10, 15, 28, 37, 38, 43, 46, 47, 63, 65, 107, 123, 125, 126, 127, 133, 139, 142, 145, 146, 150, 153, 238, 259
Data Protection Act (DPA) 2018…….……38, 125, 127, 131, 133, 134, 136, 138

TABLE OF TREATIES

European Convention on Human Rights (ECHR)……..…..…38, 131, 137, 138, 139
EU Charter for Fundamental Rights (EUCFR)…….…….……131

ACKNOWLEDGMENTS

I am very lucky to have benefited from the invaluable insight and support of my two very trusting supervisors, Professors Jane Kaye and Bettina Lange. I have also benefited a lot from being a member of two very inspiring research environments, the Centre for Health, Law and Emerging Technologies (HeLEX) and the Centre for Socio-Legal Studies (CSLS), both at the University of Oxford. In undertaking the DPhil at Oxford and writing this thesis, I have been significantly supported by generous grants from ESRC, the Onassis and the Leventis foundations. I would never have made it thus far without the constant support and love of my family and Margarita, to whom I owe my good health and sheer optimism for life.

1. INTRODUCTION

Long before the emergence of the coronavirus pandemic in early 2020, the UK government had embraced the spirit of digital reform,[1] aspiring to make online and remote civil service provision the new normal, if not the default in the long-term. To support these aspirations, a series of policy initiatives was developed, reflected in a series of high-profile reports and legislative instruments,[2] to make better use of UK government administrative data.[3] UK government administrative data is the data generated or collected throughout the operation of the UK bureaucracy, e.g. from the operation of administrative systems for such purposes as social security and tax collection.[4]

Administrative data research, i.e. research applying advanced statistical methodologies to government datasets, has been seen by policy-makers as highly promising in that regard. In the light of significant breakthroughs in big data analytics and AI over the last few decades, the UK government has sought to explore the potential of scientific analysis using 'big' government datasets to revolutionise contemporary policy-making and implementation.[5] This policy interest of the UK government has, particularly in the last few years, materialised in the introduction of numerous public-sector investments and legislative instruments to facilitate

[1] J Tomlinson, *Justice in the Digital State: Assessing the Next Revolution in Administrative Justice* (Bristol Policy Press 2019) 2.

[2] In terms of reports e.g. R Thomas and M Walport, `Data Sharing Review Report' (2008) <https://webarchive.nationalarchives.gov.uk/+/http:/www.justice.gov.uk/docs/data-sharing-review.pdf> [5.24]; Law Commission, `Data Sharing Between Public Bodies: A Scoping Report' (2014) <http://www.lawcom.gov.uk/app/uploads/2015/03/lc351_data-sharing.pdf> [7.36]; legislative instruments include the Statistics and Registration Service Act (SRSA) 2007 and the Digital Economy Act (DEA) 2017.

[3] More broadly, 'administrative' data is 'data generated during the course of some operation, and then retained in a database', see D J Hand, 'Statistical challenges of administrative and transaction data' (2018) 181(3) Journal of the Royal Statistical Society Series A (Statistics in Society) 1.

[4] R Connelly et al, 'The role of administrative data in the big data revolution in social science research' (2016) 59 Social Science Research 1, 3. There is an asymmetry in knowledge between administrative systems producing health data that can be shared for medical research, where a rich literature exists, and other types of admin data for social research. This thesis is confined to the latter.

[5] OECD, 'The Path to Becoming a Data-Driven Public Sector' <https://www.oecd-ilibrary.org/governance/the-path-to-becoming-a-data-driven-public-sector_059814a7-en> 2.

government administrative data sharing for research purposes.[6] Yet, there seems to be some distance between the policy drive and the legislative interventions to widen access to administrative data for external researchers and the actual, on-the ground experience of data sharing in the UK public sector. Empirical studies identify the existence of 'legal' and 'cultural' barriers to administrative data sharing for research purposes, raising questions about the forces that govern data access decisions.[7]

Seeking to study the regulation of research data access decisions in the UK public sector from a socio-legal perspective,[8] this thesis asks: what are the 'rules of the game' that filter access to administrative data for research purposes in the UK public sector? How and by which actors have these rules been developed over time? This chapter introduces the research problem of understanding how the conduct of data sharing for research purposes has been and is regulated in the UK over the last four decades. After a brief overview of debates in the data sharing academic and policy literature, the introduction states the argument advanced in this thesis and its contribution to existing knowledge, followed by an outline of thesis chapters.

A. Regulating Administrative Data Sharing for Research Purposes in the UK: Theory and Practice

What is administrative data sharing for research purposes and why has there been any interest to regulate it as a conduct? As mentioned, this thesis is confined to public-sector administrative

[6] See (n 2) on the legislative instruments; some examples of public-sector investments are the multi-million ESRC Big Data Network <https://esrc.ukri.org/research/our-research/big-data-network/> and the Data Science Campus <https://www.ons.gov.uk/aboutus/whatwedo/datasciencecampus> established by the Office for National Statistics in England.

[7] See the reports cited in (n 2); more broadly on legal and cultural barriers to data sharing, see Centre for Excellence for Information Sharing, 'Overcoming cultural barriers to information sharing within regulatory services' (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535439/overcoming-cultural-barriers.pdf>; G Laurie and L Stevens, 'Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom' (2016) 43 Journal of Law & Society 360, 364 also provide a long list of references on the centrality of legal barriers to data sharing, predominantly in the public health sector.

[8] By 'socio-legal' I refer to research about 'law and legal processes from a variety of social scientific standpoints', enquiring into the nature and role of law in society; see R Banakar and M Travers (eds), *Theory and Method in Socio-Legal Research* (Hart 2005) ix.

data generated throughout the operation of large-scale bureaucracies for such purposes as social security and tax collection.[9] It particularly focuses on the activity of *sharing* such data for *social-scientific research* purposes in the UK public sector. A decision to *share* data refers here to the 'provision of access to administrative datasets that contain personal data by public bodies to other entities or individuals'.[10] While not all administrative datasets contain personal data, this thesis focuses on datasets which contain personal data due to the stark difference between such datasets and more aggregate anonymised datasets that are regularly published by public bodies.[11] The latter datasets are more openly published to provide high-level information about the implementation of policies, whereas access to the former datasets can provide grounds for connecting government action with specific real-world cases, individuals and groups. Hence, there are different considerations and arrangements which make access to datasets which contain personal data much more challenging and, essentially, distinct as an activity. As far as *research* is concerned, this work confines itself to data disclosure requests coming from independent researchers, i.e. individuals or groups that are formally associated with Universities or other non-profit research centres and organisations.

Decisions to share administrative data have risen from the status of a mundane, bureaucratical task of simply managing a database, gradually emerging as a distinct area of interest that became subject to regulatory appetites. Both the government and researchers, as well as civil society actors, have begun to realise the stakes that are involved in the management of administrative data. More specifically, to begin with the UK government, recent years have seen senior government officials announcing the investment of significant resources with a view to harness the untapped potential of massive administrative datasets. This potential was expected to contribute significantly to 'public service modernisation' and improve civil service delivery through insights produced by data-intensive research.[12] The promise of such 'big'

[9] Connelly (n 4) 3.

[10] ICO, 'Data Sharing Code of Practice' <https://ico.org.uk/media/fororganisations/documents/1068/data_sharing_code_of_practice.pdf> accessed 25 July 2021 [N.B: the same date applies to all hyperlinks hereafter] – 'sharing' and 'granting/providing access' will be used interchangeably in this thesis to refer to the action of making available government data to researchers for their (research) use.

[11] For example, see the 'Destinations of key stage 4 and 16-18 pupils' and 'Pupil absence' datasets published by the Department for Education <https://explore-education-statistics.service.gov.uk/find-statistics>.

[12] J Manzoni, 'Big data in government: the challenges and opportunities' (21 February 2017)

datasets to revolutionise our empirical understanding of complex social and economic problems has fuelled this ambition.[13]

Part of this digitalisation strategy was a £64 million investment by the British Economic and Social Research Council, via the Department for Business, Energy and Industrial Strategy (BEIS), in the creation of a UK-wide infrastructure that would facilitate secure access to administrative data for research purposes.[14] The infrastructure, called the Administrative Data Research Network (the ADRN), operated between 2013 and 2018, demonstrating notable achievements, but also facing significant difficulties in facilitating access to administrative data for researchers.[15] In fact, in its fourth year of operation, the Network had managed to get the data for only ca. 19% of the ADRN-approved projects it had been facilitating.[16]

Academics and civil society organisations have also started appreciating the significance of greater access to government administrative data for research purposes. The increasing digitalisation of the civil service has made clear the significance of acquiring and maintaining more evidence on the basis of which we will be able to properly evaluate digital reforms, such as the HM Courts and Tribunals (HMCTS) £1.2bn digital modernisation investment.[17] If we look at the handling of the coronavirus pandemic as an example, the UK government declined in April 2020 to release information about the involvement of private companies like Google, Palantir and Faculty AI in the National Health Service (NHS) 'data store' which was built to help contain the outbreak. It took two months of negotiation and the threat of litigation by civil society organisations Foxglove and Open Democracy to have the contracts between the NHS

<https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities>. Also see E Dove, *Regulatory Stewardship of Health Research: Navigating Participant Protection and Research Promotion* (Edward Elgar 2020) 2 on how Research Ethics Committees (RECs) reviewing health research applications work 'under the umbrella of regulation that is becoming more streamlined and research-promoting' in England and Scotland.

[13] L Einav and J Levin, 'The Data Revolution and Economic Analysis' (2014) 14(1) Innovation Policy and the Economy 1, 7- 8.

[14] ADRN, 'History' <https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/adrn/>.

[15] Chapter 4 provides detailed discussion of the establishment and operation of the ADRN, as well as the historical development of the overarching administrative data sharing landscape in the UK.

[16] UKSA, 'Thirteenth Meeting of the ADRN Board: Agenda and Papers' (2017) <https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/06/ADRN-05-06-2017.pdf> 14.

[17] Tomlinson (n 1) 37.

and private companies disclosed.[18] Considering the heavy involvement of private companies in the digitalisation and, in some cases, the automation of civil services like benefit provision,[19] it is likely that tensions around data access for research purposes will start arising. Administrative data researchers, often affiliated with academic organisations, are both interested in and well-placed to critically scrutinise government policies that affect both individual rights and collective goods. To do so, however, they will need access to data and such access is, as it seems, far from self-evident.

These tensions and the significant attention that research data sharing decisions have attracted in the last years led to such decisions becoming subject to legal regulation and policy decisions.[20] Access decisions are no longer a matter of internal discussions by *data custodians*, i.e. public-sector organisations 'having physical and legal custody of data and deciding what is collected, aggregated and released'.[21] They are, now, a matter of extensive external debate. In the UK, data sharing was originally conceived of as an issue of public interest in the context of sharing confidential patient information for medical research purposes.[22] Since the beginning of the 21st century, however, the Department of Health has made the increased and more streamlined research use of NHS, primary care or social care data, a central aim of its

[18] Open Democracy, 'Under pressure, UK government releases NHS COVID data deals with big tech' <https://www.opendemocracy.net/en/ournhs/under-pressure-uk-government-releases-nhs-covid-data-deals-big-tech/>.

[19] UN Special Rapporteur on Extreme Poverty and Human Rights, 'Digital technology, social protection and human rights: Report' (2019) <https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx>.

[20] Although not a central concern of the present thesis, the rise of a culture of recognizing privacy rights to protect individual citizens' interests should also not be underestimated as a contributing force to making research data sharing decisions subject to legal regulation.

[21] M Burdon et al, 'Access Control in Federated Databases: How Legal Issues Shape Security' in M Wimmer et al (eds), *Electronic Government: 6th International Conference, EGOV 2007* (Springer 2007) 235.

[22] Note that the terms 'public interest' and 'public benefit' are often used interchangeably by participants in this space, either in written publications or in interview data presented in this thesis. The two terms, however, might not necessarily have the same meaning, see A Dawkins, 'Public good(s), public benefit and the public interest: A critical examination of three public purposes of English universities' in T Strike et al (eds), *Governing Higher Education Today* (Routledge 2019) 219. The thesis prefers the notion 'public interest' and uses it for consistency; the use of 'public benefit' by some of my interviewees shall be read as interchangeable with the 'public interest'.

research governance framework.[23] Legislative acts, such as the National Health Service Act (NHS Act) 2006 and the Health and Social Care Act (HSCA) 2012 delineate the powers of custodians to enable health data sharing for research purposes.[24]

More recently, policy interest was extended to non-health administrative data and its potential use by social researchers. A series of White Papers, independent reviews and reports produced or commissioned by the UK Government suggested that the broader use of public-sector data to facilitate research and policy development should be regarded as a priority.[25] Legal developments followed in the interest of providing a 'robust UK-wide evidence base to inform research, thereby guiding the development, implementation and evaluation of policy'.[26] The Digital Economy Act (DEA) 2017 established a generic legal power for public-sector custodians to disclose data they hold 'in the context of their functions' (s. 64 DEA) for research purposes, thus enhancing existing data sharing legal powers.

Based on these developments, does the law enable wider access to administrative data for research purposes? This question has been in the minds of policy-makers in the UK for some time. In 2015, the Law Commission enquired into:

> 'whether there are inappropriate *legal or other hurdles* to the transfer of information between public bodies and, potentially, between public bodies and private bodies engaged in public service delivery' (emphasis added)[27]

---

[23] Department of Health, 'Research Governance Framework for Health and Social Care' (2nd edn, 2005) 17; National Institute for Social Care and Health Research Wales, 'Maximising the Use of Routine Data for Research in Wales' (2013) <https://www.healthandcareresearch.gov.wales/uploads/Policy%20%26%20Strategy/Use%20of%20routine%20data/Maximising%20the%20Use%20of%20Routine%20Data%20for%20Research%20in%20Wales.pdf> 3.13: 'Maximising the use of routine data can play a key role in enhancing the evidence base and offering innovative approaches to policy development and evaluation'.

[24] Section 251 of the National Health Service Act 2006 and section 259 of the Health and Social Care Act 2012 are relevant here.

[25] Cabinet Office, 'Open Data: unleashing the potential' (28 June 2012); S Shakespeare, 'Shakespeare Review: An Independent Review of Public Sector Information' (May 2013); Administrative Data Taskforce, 'The UK Administrative Data Research Network: Improving Access for Research and Policy' (December 2012).

[26] Administrative Data Taskforce *ibid* 2.

[27] Law Commission (n 2) 1.2.

The Commission identified relevant hurdles by reference to a series of empirical studies,[28] as well as other academic and policy studies of risk aversion in UK public-sector data sharing.[29] This was found to create uncertainty about the actual exercise of legal powers to share administrative data in general.[30] A prime example of this uncertainty is an independent review report of the ADRN investment, which struggled to explain the relatively more cautious exercise of data sharing powers in England compared to the other UK countries.[31] This report identified a 'lack of legal clarity about the ability of government departments or agencies to provide data to the ADRN' as one of the main obstacles that the Network had to tackle before accessing government data.[32]

Along the same lines, the Law Commission's conclusion as to how relevant barriers to information sharing are to be addressed was that legal reform can 'play a part in addressing problems of data sharing between public bodies'.[33] Nonetheless, the report recorded a *variety* of opinions in respect of how exactly this could happen. The Law Commission itself did not advocate for wider legal powers but for more clarity, also stressing that 'the problem is not one

[28] House of Commons, 'The Bichard Inquiry Report' (2004) <https://dera.ioe.ac.uk/6394/1/report.pdf>; M Flynn, 'South Gloucestershire Safeguarding Adults Board, Winterbourne View Hospital, A Serious Case Review' (2012) <https://hosted.southglos.gov.uk/wv/report.pdf>; Birmingham Safeguarding Children Board, 'Serious Case Review in respect of the death of Keanu Williams' (2013) <https://www.cscb-new.co.uk/downloads/Serious%20Case%20Reviews%20%20exec.%20summaries/SCR_Archive/Birmingham%20SCR%20-%20KW%20(2013).pdf>.

[29] Thomas and Walport (n 2) 5.24; Laurie and Stevens (n 7) 362.

[30] Note that legal powers in this context are always discretionary, with the UK government finding that 'mandating the disclosure of data would not be appropriate', Department for Culture, Media & Sport, 'Digital Economy Bill – Factsheet – Digital Government: Research and Statistics (clauses 56-68)' <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535021/11._Research_and_Statistics.pdf>.

[31] Research Project Management, 'Administrative Data Research Network (ADRN) Mid-Term Review Report' (8 November 2016) <https://esrc.ukri.org/files/research/research-and-impact-evaluation/adrn-mid-term-review-report-8-november-2016/> 3, 7; a similar analysis is presented in K H Jones et al, 'The Good, the Bad, the Clunky: Addressing Challenges in Using Administrative Data for Research' (2019) 4(1) International Journal of Data Population Science.

[32] Research Project Management *ibid* 7.

[33] Law Commission (n 2) 11.1.

of lack of power to share so much as disinclination to use the powers'.[34] The Information Commissioner's Office held that 'a more prominent role for data protection law' would be critical to avoid confusion for practitioners.[35] Within the same report, Government departments and public bodies called for a codification of existing law to provide them with a 'map or single table of existing legislative data sharing powers'.[36]

While views echoing the centrality of legal hurdles to data sharing were not voiced for the first time,[37] the Law Commission report also recorded a different view on the matter. This view, also defended by the academic work of Laurie and Stevens, highlights the centrality of an organisational *culture* that 'influences how personal information is viewed and handled'. This culture creates barriers for research data sharing by aggravating 'perceptions by all parties of risk that will come from that activity'.[38] Authoritative policy reports written by senior UK government officials,[39] which hold the same view, concur that a culture shift within the UK public sector is more important than any legal amendment to alleviate data sharing barriers.

Against this backdrop, the central aim of this thesis is to explore from a socio-legal perspective 'what is going on' in the space of research data sharing decisions. While the focus of this thesis remains on the administrative data sharing context, whose particularities are not to be underestimated with regard to this work's findings, the research problem can be conceived to have broader implications: who regulates the employment of new technologies,[40] like data sharing, in the public sector? What are the forces that structure decision-making and drive the exercise of regulatory power towards particular outcomes? Does the 'law' primarily shape decisions to disclose data in the UK public sector, or do other forces steer such decisions

[34] *Ibid* 11.4.

[35] *Ibid* 11.5.

[36] *Ibid* 11.62.

[37] Administrative Data Taskforce (n 25) 16; Laurie and Stevens (n 7) 377 citing academic literature on data sharing in the public health sector.

[38] Laurie and Stevens *ibid* 362.

[39] Thomas and Walport (n 2) 5.24; also see Centre for Excellence in Information Sharing (n 7) 7.

[40] Administrative data sharing is a 'new technology' to the extent that hitherto unprecedented volumes of government data have started to be generated, as well as ground-breaking new technologies of data analytics have become possible, in the last few decades.

to a greater extent than the law? The following section presents a high-level view of the key findings of this study.

B. Theorising a Research Data Sharing Decision-Making Space: Unwritten Norms and Informal Interactions

Organisational actors engage in research data sharing relationships within a particular 'space', demarcated, as it will be shown, by the main considerations that shape the process of data access decision-making and enabling particular types of research being granted access.

Working within the 'regulatory space' analytical framework, this thesis has identified and examined three dimensions of the 'research data sharing space': the *actors*[41] which have populated the space and have been involved in shaping the data sharing frameworks over time,[42] the *institutions* or 'rules of the game'[43] that govern research access to data, and the *interactions* through which the actors form their organisational identities by reference to the institutions, as well as create the institutions themselves. In this space, as it will be shown, actors draw references from the 'rules of the game' that filter access to data to construct their identities, e.g. as pioneers of data-driven research and facilitators of research on government data, while at the same time leveraging their resources and position in the space to confine administrative data research to particular, non-threatening from their standpoint types of research.[44] In studying the development of this space, the thesis adopted an in-depth, qualitative research design, which employed semi-structured interview and documentary data about the regulation of research data sharing decisions in the UK public sector.

---

[41] While the analysis in this thesis often includes the contribution of particular individuals to research data sharing regulation, 'actors' shall be taken to primarily refer to organisational actors.

[42] 'Data sharing frameworks' refer to the organisational research data sharing policies, procedures and practices that filter access to data in practice. According to the argument advanced in this thesis, as it will be outlined in this section, they are shaped predominantly by specific research data sharing institutions or 'rules of the game'.

[43] On the definition of institutions, see *infra* chapter 2, C.

[44] The Department for Work and Pensions (DWP) is a prime example here, drawing inspiration from the UK Government's broader policies to bolster its reputation as a body which promotes wider use of administrative data for research, while, on the other hand, adopting a non-facilitative to independent researchers organisational policy towards research data sharing, see following chapters on this.

The argument made in this thesis can be summarised in the following way: in the UK public sector, research data access has been governed by rules which are incrementally established and revised via small-scale interactions between low-level technical staff, rather than momentous legal and policy decisions made by high-level officials or bodies. These commonplace interactions, however, conceal, behind their routine and technical character, underlying political rationales which stem from the organisational history and mission of each public body. Through the commonplace interactions, unwritten, conventional normative assumptions about the ownership of administrative data and the relationship of administrative data research with policy-making determine the types of data sharing frameworks that are established and implemented. In that sense, the 'political' is artfully concealed behind the 'technical' and the 'commonplace'.

This argument contributes both to the nascent body of public-sector data sharing literature and to the broader regulatory space literature. Firstly, it contributes to the emerging literature on administrative data sharing for research purposes in the public sector.[45] It does so by developing an analytical and methodological framework to assess the validity of existing views in the literature about the relative strength of legal or other obstacles to administrative data sharing for research purposes.[46] Considering that most of the literature on administrative data sharing is not academic but comes from policy organisations, the present research design provides an in-depth understanding about the role of law in public-sector data sharing regulation. Such an understanding is both significantly broader and deeper than existing literature accounts: while these accounts offer us with a number of general-level potential influences (e.g. 'law' or 'culture') on research data access decisions, the present thesis undertakes an empirically-informed assessment of the way in which these influences have manifested themselves in specific organisational environments across the UK public sector. In doing so, it draws on both primary and secondary qualitative data that relates both to the historical development of research data sharing regulation and its contemporary dynamics. The understanding offered by this thesis is helpful not only from a theoretical perspective, but also

[45] M Iveson and I Deary, 'Navigating the landscape of administrative data in Scotland' (2019) 4 Wellcome Open Research 97; Law Commission (n 2); Research Project Management (n 31); Thomas and Walport (n 2); Laurie and Stevens (n 7).

[46] Law Commission *ibid; cf* Laurie and Stevens *ibid*.

to practically assess the extent to which legal strategies such as litigation are likely to yield significant results for research data access purposes.

Secondly, this thesis contributes to regulatory space studies particularly with regard to assessing the relevance of unwritten normative frames and commonplace, low-level interactions for regulatory regimes. It does so by elaborating on the nuanced ways in which a collective 'research data sharing culture' shapes the formation of governance frameworks in a particular space and, thus, the exercise of regulatory power in that space.

While regulatory space approaches see 'law' as only one of the resources available to actors,[47] they often analyse power struggles by reference to an actor's 'capacity to dominate the interpretation of regulatory obligations'.[48] This reading associates regulatory obligations with codified and formally articulated requirements.[49] Contributing to such studies, the present argument aspires to enrich theoretical understanding of the ways in which relationships, often informal and small-scale, between organisational actors shape who is 'in' and who is 'out' of a regulatory space. The creation, revision or repeal of research data sharing frameworks is explained by reference to alliances and struggles between actors in the space, by reference to the formation of 'rules of the game' that are often neither codified nor formally articulated. This thesis seeks to capture subjects' perceptions of the 'law' and other regulatory forces, such as cultures of administrative data ownership and collective beliefs about the value of policy-relevant research. To do so, perspectives from individuals who operate within smaller or larger government and other broader public-sector organisations are examined, diverging from the high-level approaches usually associated with studying the dynamics of a whole space. This introductory chapter concludes with an outline of the chapters included in the present thesis.

[47] L Hancher and M Moran, 'Organizing Regulatory Space' in R Baldwin, C Scott and C Hood (eds), *A Reader on Regulation* (OUP 1998) 150; C Scott, 'Analysing Regulatory Space: Fragmented Resources and Institutional Design' (2001) Public Law 283, 287.

[48] Scott *ibid* 285.

[49] M Canning and B O'Dwyer, 'The Dynamics of a Regulatory Space Realignment: Strategic Responses in a Local Context' (2013) 38 Account Org Soc 169-70; K Jonnergård and U Larsson-Olaison, 'Developing Codes of Conduct: Regulatory Conversations as Means for Detecting Institutional Change' (2007) 29(4) Law & Policy 460.

C. Outline of Thesis Chapters

The present thesis is comprised of eight (8) chapters. Following this introductory chapter, chapters two (2) and three (3) present the analytical and methodological framework within which this enquiry has been undertaken. Chapters four (4) to seven (7) present the findings of the thesis, drawing on primary and secondary qualitative data. Chapter eight (8) concludes the thesis.

Chapter two (2) presents the analytical framework of the thesis, proposing a lens that draws on the regulation theory concept of 'regulatory space'. The chapter begins by clarifying the particular relevance of this concept for studying the regulation of the 'process' rather than the regulation of the 'outcomes' of data access decisions. Regulatory space approaches, by focusing on the interdependence between organisational actors and their conflict and bargain around an array of regulatory issues, are apt for explaining process-related questions about regulation. Regulation of access decisions is examined in this thesis under three analytical dimensions of the 'research data sharing space': the 'actors' which have populated the space and have been involved in shaping data sharing processes and frameworks over time; the 'institutions' or 'rules of the game' that filter access to administrative data for research purposes and the 'interactions' between the actors, through which the space has been shaped and recalibrated over time, e.g. at times 'opening up' to allow wider access for independent researchers and at other times 'closing down' to confine data access to researchers who can demonstrate a clear policy relevance. The theoretical categories emerging from analysing the data collected in this thesis have been organised within the three dimensions of actors, institutions and interactions.

Chapter three (3) presents the methodological framework developed for the thesis: an in-depth, qualitative research design, which employed semi-structured interview (involving 42 participants) and documentary data about the regulation of research data sharing decisions in the UK public sector. The chapter begins by providing a micro-level 'anatomy' of research data sharing decisions, fleshing out the key roles and job types of individuals involved in this activity to clarify the focus of the empirical enquiry. This work's approach to data collection and analysis is then discussed, drawing on the particularities of the population of UK research data sharing participants. The strategic choices made during the different phases of this work, as well as the sampling of research participants are explained and justified. Finally, this work's

efforts to enhance the 'validity' and 'reliability' of its findings are presented, including the design of a small-scale, pilot survey. The chapter concludes with a discussion of this work's methodological limitations.

Chapter four (4) considers the way in which the actors populating the research data sharing space in the UK have shaped its historical development over the last four decades. Under the lens of regulatory space, the chapter sketches the way in which particular organisational actors have defined the scope of central issues to the development of the space over the last four decades. Two central narratives invoked by actors stand out and determine who is 'in' and who is 'out' of the research data sharing space: 'efficiency', referring to data-intensive research as a way to improve policy-making, and 'accountability', referring to data-intensive research as a method of critically scrutinising the government's policies in the name of transparency and openness. I discuss how interactions between the actors marked transitions among the three main historical phases of research data sharing regulation, showing that certain types of independent research have been, at times, included and, at other times, excluded from the space. Based on the analysis of the historical development of the space, the chapter concludes with the main questions about the formation of research data sharing institutions in the UK public sector, to be addressed in the following chapters.

Chapter five (5) shifts analytical attention to the 'rules of the game' that filter access to administrative data and, more specifically, to the interpretation of 'legal' powers and limits and their relevance for research data sharing regulation. This chapter enquires into whether legal considerations related to research data sharing determine the creation and application of data sharing frameworks in practice. It proceeds from an analysis of the formal legal framework governing research data sharing in the UK, extended by the understandings of 'legal' powers and limits shared by the participants of this study. Then, the chapter presents this study's empirical findings, according to which 'legal' considerations are not the most determinative factor of the way in which research data sharing decisions are regulated in the UK.

Chapter six (6) asks: if legal considerations are not predominantly shaping the 'rules of the game' that filter access to administrative data for research purposes, where do these rules come from? Drawing on this work's empirical findings, the chapter unravels the existence of a collective 'research data sharing culture' across the UK public sector. The two core constituents of this culture are unpacked: unwritten, uncodified and conventional normative assumptions

about the ownership of administrative data and the relevance of research to policy-making priorities. Particular views as to the ownership of administrative data, or as to the desired alignment between independent research and a custodian's policy aims, are associated, as the chapters shows, with the creation and implementation of particular types of data sharing frameworks.[50]

Chapter seven (7) addresses the last, yet crucial dimension of the space constituted by the taking of data access decisions: how do the interactions between the actors within this space shape the regulation of research data sharing decisions? The chapter argues that commonplace, routine and low-level communications, often between data analysts, are more crucial than high-level and momentous legal and policy decisions in determining the scope and content of research data sharing frameworks. Disagreements, concerns and crises are negotiated and perceptions about the desirability of administrative data research are shifted through informal, small-scale iterations, not formal decision-making structures. Either with regard to 'opening up' the boundaries of the space to widen access to administrative data for independent researchers, or in 'closing down' the same boundaries to confine data access to certain types of research, such small-scale interactions matter. The chapter proceeds from unpacking the role of data analysts as the prominent network behind these interactions, before discussing the nature of the interactions themselves as informal and incremental, i.e. likely to result in bigger changes through small steps.

Chapter eight (8) draws the core tenets of the analysis together. A summary of the theoretical and methodological contributions of the thesis to regulatory space and data sharing studies is provided. The UK public-sector research data sharing space is a space within which commonplace, informal and small-scale interactions, taking place within inter-organisational and cross-organisational relationships, define the scope of regulatory issues and recalibrate the boundaries of the space over time. Such relationships, which are often formed and maintained far from the policy-making spotlight, merit analytical attention as much as momentous legal and policy decisions made by high-level officials or bodies do. On a more normative level, recommendations are made in the interest of facilitating disclosure-friendly data sharing frameworks and, thus, promote academic, data-driven research in the public interest. Particular communicative strategies, it is suggested, hold significant potential to lay the foundations for

[50] See (n 42) on the definition of such frameworks.

good relationships between data custodians and researchers. In the shorter term, this will facilitate research data sharing, while instigating, in the longer term, a reasonable shift in deeply-held organisational values about public-sector data. Reflections on the legal and cultural prerequisites for the success of such strategies, as well as directions for further interdisciplinary research into the regulation of research data sharing, are offered.

2. THEORISING THE 'DECISION-MAKING SPACE' OF ADMINISTRATIVE DATA SHARING FOR RESEARCH IN THE UK: ACTORS, INSTITUTIONS AND INTERACTIONS

This chapter presents the analytical framework of the thesis, proposing a lens that draws on the concept of 'regulatory space'. As noted in the introduction,[51] this thesis seeks to explain how research data sharing decisions are and have been regulated in the UK public sector over time. I begin this chapter by clarifying that my focus is an analysis of the regulation of the *process* rather than the outcomes of data custodian decision-making. I then introduce the concept of 'regulatory space'.

After that, I present the three dimensions of this analytical framework: the *actors* which populate and have populated the 'research data sharing space' over the last four decades; the *institutions* which I define here as 'rules of the game' that filter access to administrative data for research purposes and the *interactions* between the actors which shape and recalibrate data sharing institutions within this space.[52] This tripartite distinction brings to the fore the particular characteristics of this space: multiple actors need to collaborate with one another over a long time, repeating routinised technical processes, such as linking and cleaning datasets, while demonstrating conformity to organisational and legal requirements. The exercise of power in this space is not to be understood only by reference to the formation and interpretation of formally articulated regulatory requirements, but also by understanding how more dynamic 'rules of the game' governing research data sharing are shaped in these commonplace and routine interactions. The concept of 'regulatory space' is apt for conducting empirical research into how organisational and individual *actors* create, preserve or modify, through their *interactions,* research data sharing *institutions,* i.e. a range of data sharing standards and criteria. The following chapter 3 will elaborate on the methodological choices of the thesis in conducting empirical research in this context, before chapters 4-7 present my findings about the regulation of research data sharing decisions in the UK.

[51] *Supra* chapter 1, A-B.

[52] See *infra* p 44 on 'institutions' as distinct from 'organisations'; throughout the present thesis, 'organisations' will refer to actors and 'institutions' will refer to norms, conventions, customs, ideals and frameworks governing research data sharing.

A. Regulating 'Research Data Sharing Decisions' from a Spatial Perspective: Actors, Institutions and Interactions

I first clarify my interest in how the *process* of data custodian decision-making is regulated in the UK public sector, before introducing the concept of 'regulatory space' as apt for addressing my research questions.

i. Regulating 'Research Data Sharing Decisions': Outcomes or Processes?

Administrative data sharing for research purposes in the UK public sector involves, as an operational activity, at least three parties: a researcher or, more commonly, a research group which is interested in accessing government data; a data custodian, [53] i.e. a public body which is responsible for managing the data and granting or rejecting access to it for all purposes and one or more trusted intermediary bodies[54] which will undertake the responsibility of performing technical tasks (e.g. data linkage or de-identification) and generally facilitating the process (e.g. by accrediting or training researchers in how to securely use data). The introduction to this thesis set out two conflicting views, i.e. one that emphasises the role of formal state law and another one that considers the role of 'organisational culture' as central to hindering the development and application of consistent data sharing procedures.[55]

This disagreement is to be understood from a *regulatory* perspective. Enquiring into the salience of law in governing a particular type of conduct that is subject to policy decisions raises questions about the centrality of law as an 'instrument of social control' within a

[53] Burdon (n 21) 235: a 'data custodian' is an entity 'having physical and legal custody of data and deciding what is collected, aggregated and released'. Data 'sharing' and data 'access' are used interchangeably in this chapter to refer to the disclosure of datasets by government data custodians to third-parties.

[54] Intermediary bodies are not interested in accessing administrative datasets themselves, but they exercise stewardship functions by alleviating uncertainty and risks around data sharing. For an analysis of the role of such bodies, see G Laurie et al, 'Charting Regulatory Stewardship in Health Research: Making the Invisible Visible' (2018) 27 Cambridge Quarterly of Healthcare Ethics 333.

[55] This is not to say that this is necessarily a clear-cut dichotomy, yet proponents of these views are: Law Commission (n 2) 1.2; Administrative Data Taskforce (n 25); *cf.* Laurie and Stevens (n 7) 362; Thomas and Walport (n 2) 5.24.

regulatory regime that seeks to give these policies effect.[56] *Regulation* is broadly defined here as establishing norms through setting rules or standards, as well as monitoring and enforcing compliance with them via both legal and non-legal processes.[57] Regulatory processes are designed to influence behaviour, including through facilitating and expressing certain 'clusters of ideas and assumptions about how things are to be done'.[58] Research data sharing is an area where public policy objectives are explicitly pursued by the State, partly through a system of purposively designed legal norms, which are not, however, always determinative of actual organisational data sharing frameworks and practices. Such frameworks and practices are a reflection of the 'rules of the game' governing research data sharing, defined in this thesis as research data sharing *institutions*. 'Institutions' are defined in organisational studies as the 'rules, procedures, organisational standards, governance structures, conventions and customs',[59] as well as the 'symbol systems, cognitive scripts and moral templates',[60] which operate within or interact with organisational environments and structure individual behaviour. Institutions include 'both formal structures and informal rules and procedures that structure conduct'.[61]

What does it mean, however, to regulate 'data sharing decisions'? There are at least three potential dimensions of regulatory interest when exploring the different 'legal' and 'non-legal' influences on research data sharing decisions in the UK public sector. First, one could focus on individual *decision-making outcomes*, i.e. 'access' vs 'non-access', by reference to the role of legal and other factors in steering them. Second, without focusing on individual outcomes, one could seek to explain the role of these factors in creating *decision-making patterns,* i.e. stable and recurrent tendencies to respond to research requests in a particular manner, granting or

[56] B Morgan and K Yeung, *An Introduction to Law and Regulation* (CUP 2012) 1.

[57] B Lange and F Haines, 'Regulatory Transformations: An Introduction' in Lange, Haines and Thomas (eds), *Regulatory Transformations: Rethinking Economy-Society Interactions* (Hart 2015) 1, 7; similarly, and even more broadly, in Dove (n 12) 3 where regulation is defined as a 'set of rules, principles, mechanisms, strategies, or activities (…) that either affect behaviour as an incidental effect or are designed to steer behaviour in a (…) desirable way'.

[58] R Baldwin, M Cave and M Lodge, *Understanding Regulation: Theory, Strategy and Practice* (OUP 1999) 3.

[59] W Powell and P Dimaggio, 'The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields' (1983) 48(2) American Sociological Review 147.

[60] P Hall and R Taylor, 'Political Science and the Three New Institutionalisms' (1996) Political Studies 936, 940.

[61] K Thelen and S Steinmo, 'Historical Institutionalism in Comparative Politics' in K Thelen, S Steinmo and F Longstreth (eds), *Structuring Politics: Historical Institutionalism in Comparative Analysis* (CUP 1992) 2.

rejecting disclosure. This would also involve to explain what shapes decision outcomes, albeit only at a higher, more aggregate level. Third, shifting analytical attention away from the actual, individual or aggregate, decision outcomes, one could seek to understand how the dominant, legal or non-legal, *processes and frameworks* governing the activity of research data sharing are constituted. The latter analytical focus, i.e. the *process* through which research data sharing frameworks are formed and become prevalent, is the one of the present study. Rather than asking what determines the outcome of particular access decisions, this type of enquiry is interested in what *structures* and *justifies* the *process* of deciding about access from the perspective of the actors in this space, mostly data custodians, researchers and intermediary bodies.[62]

There are two reasons for choosing this focus. First, while the first two foci are arguably compelling proxies of how the law and other factors influence custodian decision-making, there is, presently, a lack of adequate information to pursue them. There is a dearth of information about actual research data access decision outcomes, with only few public bodies in the UK publishing relevant information[63] and the majority being resistant to publicise concrete data.[64] As a result, it is not possible to state, even by approximating the number of data custodian decision-makers,[65] or the number of requests particular bodies have received, granted or rejected. Second, normative frameworks and organisational procedures matter regardless of their causal relationship to modifications in participant behaviour. As a representation of organisational efforts to 'enact order and coordinate action',[66] such frameworks and processes communicate both to internal and external actors an organisation's

[62] See *infra* p 33 on who are these actors and the relationships between them.

[63] The Department for Education (DfE) in England is a luminary in respect of publishing data on third-party requests and their data sharing decisions with regard to the 'National Pupil Database', Department for Education, 'DfE external data shares' (18 July 2019) <https://gov.uk/government/publications/dfe-external-data-shares>.

[64] R Booth, 'Benefits system automation could plunge claimants deeper into poverty' (The Guardian) 14 October 2019 <https://www.theguardian.com/technology/2019/oct/14/fears-rise-in-benefits-system-automation-could-plunge-claimants-deeper-into-poverty>.

[65] *Cf.* for example the website of the Dutch General Administrative Service <https://www.algemenebestuursdienst.nl/>, which hosts an online database with the names and roles of high-ranking Dutch civil servants.

[66] W Powell and J Colyvas, 'Microfoundations of Institutional Theory' in R Greenwood et al, *The Sage Handbook of Organizational Institutionalism* (Sage 2008) 282.

purposes, rationality and wider beliefs and values in respect of a particular activity.[67] At the least, such communications provide fruitful insights into how organisations perceive themselves, their responsibilities and their roles in society, with scholars arguing that they have such significant functions as 'defining available modes of action', 'constraining' and 'empowering' behaviour.[68]

That being said, seeking to understand how the frameworks and processes governing the activity of research data sharing are being formed and applied is associated with its own challenges. Research data sharing has been developing in an incremental manner, with departments reactively responding to requests for sharing data rather than pro-actively developing their own procedures.[69] Even after the introduction of a generic legal power for public bodies to share data for research purposes in the UK Digital Economy Act (DEA) 2017,[70] there is not a public sector 'gold standard' on the appropriate decision-making actors and procedures or the relevant considerations that should govern data access decisions.[71] Instead, public bodies are expected to develop their own standards and procedures, ensuring that these conform to such broad principles, enshrined in the Code of Practice published by the UK statistics authority, as 'proportionality', 'transparency' and 'data minimisation'.[72]

With a fair amount of uncertainty about the prevailing frameworks and processes being voiced,[73] it will be for this thesis to use existing secondary and primary original data collected by myself, in order to shed light on the 'rules of the game' that govern this form of decision-

[67] R Scott, *Institutions and Organizations* (Sage 1995) 134.

[68] *Ibid*: ' Institutions construct actors and define their available modes of action; they constrain behaviour, but they also empower it'.

[69] M J Taylor et al, 'Developing a data sharing framework: a case study' (2014) 8 Transforming Government: People, Process and Policy 151.

[70] Section 64 of the Digital Economy Act 2017 established a generic legal power for public-sector custodians to disclose data they hold 'in the context of their functions'.

[71] National Audit Office, 'Challenges in using data across government' (21 June 2019) <https://nao.org.uk/wp-content/uploads/2019/06/Challenges-in-using-data-across-government.pdf> 11.

[72] UK Statistics Authority, 'Research Code of Practice and Accreditation Criteria' (1 March 2018).

[73] Iveson and Deary (n 45) 97; F Ritchie, 'UK Official Microdata Release Practices' <https://www.academia.edu/1336375/UK_Official_Microdata_Release_Practices> 3.

making in the UK public sector.[74] The following section introduces the concept of 'regulatory space' and its application through three different dimensions, i.e. actors, institutions and interactions. This thesis will be developing arguments about the process of research data sharing decisions within these three dimensions.

ii. The 'Space' of Making Decisions about Research Data Sharing and its Dimensions

'Regulatory space' is an appropriate concept for understanding the process of making decisions about research data sharing, where multiple organisations enter complex and long-standing relationships within which data sharing standards are (re)negotiated and conformity to them is demanded from more than one parties. Originally developed by Hancher and Moran as a theoretical construct for understanding economic regulation by the State under advanced capitalism, 'regulatory space' analysis sees regulation as a 'process of intermediation and bargaining between large and powerful organisations'.[75] A 'regulatory space' is defined as an:

> 'abstract conceptual space constructed by people, organisations and events acting in concert upon a set of specific regulatory issues subject to public decision'.[76]

Regulatory space analysis challenges conventional divides between 'public' and 'private' theories of regulation, which see regulation either as State intervention in the interests of a wider community or as informed by private interests.[77] Regulation is, thus, not understood in terms of 'discrete, binary, hierarchical relationships between regulators and regulated'[78] but as an activity shaped by the *interdependence* of 'large, sophisticated and administratively complex organisations performing wide-ranging economic and social tasks'.[79] Within the

[74] O Williamson, 'Comparative Economic Organization: The Analysis of Discrete Structural Alternatives' (1991) 36 Administrative Science Quarterly 269 including 'customs, laws and politics' within the 'institutional rules of the game'.

[75] Hancher and Moran (n 47) 150.

[76] J Young, 'Outlining regulatory space: Agenda issues and the FASB' (1994) 19(1) Accounting, Organizations and Society 83, 84.

[77] Morgan and Yeung (n 56) 16.

[78] B Lange and C Cook, 'Mapping a Developing Governance Space: Managing Drought in the UK' (2015) 68 Current Legal Problems 229, 252.

[79] Hancher and Moran (n 47) 148.

regulatory space, roles between 'rule makers, enforcers and bearers of sectional interests' are not static, but constantly shift,[80] according to changes in the dispersion of regulatory authority. Such authority is fragmented among a 'number of organisations, public and private', with legal authority being only one among several sources of power.[81]

While the concept of 'regulatory space' was devised to theorise about business regulation and interrogate the balance of power between the State and powerful *private* actors, it has been employed to explain the dynamics of regulatory processes more generally. As Scott observes, the concept can be used to provide an empirically-grounded understanding of the 'configuration of resources and relations within the regulatory space'.[82] The present thesis will be using the concept in this sense and, for present purposes, the role of market actors, such as data-savvy technology companies, in regulating research data sharing decisions will be set aside. This is not to say that this is a less important question, with both legal definitions of data-intensive research on a European Union level[83] and policy agendas in the UK[84] endorsing industry access to administrative data for research purposes. Yet, public-private research data sharing partnerships are relatively scarce in so-far organisational practice in the UK[85] and interrogating their dynamics would require a greater level of maturity of empirical experience.

The type of data access decisions I focus on in this thesis are embedded in long-term relationships, involving various stages of communication and negotiation, which are developed between data custodians, intermediary bodies and researchers. Due to its nature as an activity,

---

[80] *Ibid* 152.

[81] Scott (n 47) 287.

[82] *Ibid* 286.

[83] Recital 159 of the Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/ (General Data Protection Regulation) [2016] OJ L119/1: 'For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including (…) privately funded research'.

[84] REFORM, 'Making NHS Data work for Everyone' <https://reform.uk/research/making-nhs-data-work-everyone>.

[85] *Cf* a collaboration between Swansea University (as 'data custodians' of the SAIL databank) and Quintiles, a major player in the healthcare industry, on managing patient engagement for research purposes across clinical studies; E Cascade et al, Combining Electronic Health Records and Patient-Reported Information in the UK: Preliminary Results from the WASPS (Wales SAIL + PRO Study' (2012) Value in Health.

research data sharing involves the co-ordination of technical processes and dynamic interactions between these stakeholders, who may have conflicting interests in accessing or withholding administrative data. These interactions may happen within either 'well drawn and institutionalised settings and through formalised processes' or 'more informal' processes, without necessarily resulting in the culmination of a formal or legally enforceable output.[86] Formal and informal means of regulation may 'exist in a symbiotic relationship',[87] with actors who lack the formal powers to shape the outcome of regulation often playing a critical role in doing so.

Hence, my analysis of how the research data sharing decisions are regulated in the UK public sector proceeds from three conceptual premises: *actors, institutions* and *interactions*. While the categories that emerged out of my empirical data analysis will be linked to these three premises, it is important to stress from the outset that this analysis is inductive, i.e. no assumptions from the conceptual framework were projected into the data. Data analysis first yielded specific empirical categories (e.g. a 'research data sharing culture' governing decisions)[88] and these categories were, in turn, grouped together under each of the three premises (e.g. 'institutions'). In other words, the three premises have been applied as meaningful analytical labels to structure the present enquiry. This is not to say that these premises are always the central ones within regulatory space studies, with other concepts like the 'boundaries' of the space and the 'resources' which are employed by the space's constituents being equally significant in scholarly accounts.[89] Nonetheless, the three dimensions I will be unpacking here were the most relevant ones to my empirical categories and my research questions.

First, organisational *actors* are the 'primary drivers of regulatory actions' within a regulatory space approach.[90] A regulatory space approach observes regulatory processes from a 'closer perspective' and rejects a 'rigid hierarchical' analysis often assumed within regulatory

[86] N Chowdury and R Wessel, 'Conceptualizing Multilevel Regulation in the EU: A Legal Translation of Multilevel Governance?' (2012) 18(3) European Law Journal 335, 339.

[87] R MacKenzie and M Martinez Lucio, 'The colonisation of employment regulation and industrial relations? Dynamics and developments over five decades of change' (2014) 55(2) Labour History 189, 193.

[88] *Infra* chapter 6 elaborates on this culture.

[89] Scott (n 47) 287.

[90] Chowdury and Wessel (n 86) 340.

studies.[91] This approach encourages empirical analysis of the roles and aims of a diverse number of actors within a specific regulatory regime.[92] The roles and aims of such actors as research groups and consortia, public-sector data custodians and research funders as intermediaries will be analysed in this thesis. Second, such actors 'participate in public policy-making' not only creating, but also being themselves constituted by,[93] 'certain *institutions* in regulation' (emphasis added).[94] As mentioned before, this thesis defines 'institutions' as the 'rules, procedures, organisational standards, governance structures, conventions and customs',[95] as well as the 'symbol systems, cognitive scripts and moral templates',[96] which operate within or interact with organisational environments and structure individual behaviour. Both formalised and informal institutions 'that structure conduct',[97] such as data sharing legal rules, collective assumptions about the value of data-driven research and perceptions of administrative data ownership, will be analysed. Third, the *interactions* and interdependencies between multiple actors will be a focal aspect of studying regulatory processes in the research data sharing space.[98] This focus will allow an understanding of how certain groups or communities in the space, e.g. organised research groups or civil servants operating as data custodians, draw on 'plural systems of authority and of other resources'[99] in creating or implementing a regulatory regime through their interactions.[100] Through their interactions, organisational actors create, modify and preserve institutions which structure and justify research data sharing in the UK public sector. The application of these three dimensions to the research data sharing space will now be analysed in turn.

[91] F Vibert, *The New Regulatory Space: Reframing Democratic Governance* (Edward Elgar 2014) 4.

[92] Lange and Cook (n 78) 251-2.

[93] R Scott (n 67): ' Institutions construct actors and define their available modes of action; they constrain behaviour, but they also empower it'.

[94] Scott (n 47) 293.

[95] Powell and Dimaggio (n 59) 147.

[96] Hall and Taylor (n 60) 940.

[97] Thelen and Steinmo (n 61) 2.

[98] J Kaye and S Gibbons, 'Mapping The Regulatory Space For Genetic Databases And Biobanks In England And Wales' (2008) 9(2) Medical Law International 111, 112.

[99] Scott (n 47) 287.

[100] J Black, 'Regulatory Conversations' (2002) 29 Journal of Law and Society 163; Jonnergård and Larsson-Olaison (n 49) 471.

B. Mapping the Actors Populating the Research Data Sharing Space

Regulatory space analysis is a variant of historical institutionalism, a research tradition which aims to enhance our understanding of complex political phenomena by reference to 'the origins, evolution, and consequences of humanly created institutions across time and place'.[101] In this tradition, historical path dependencies and context-specific justifications explain the prevalence of particular social structures and organisational procedures which govern and constrain individual behaviour.[102] In the interest of providing a comprehensive understanding of the actors and dynamics of the particular space, chapter 4 will be drawing on documentary data, such as legal and policy documents enacting research data sharing frameworks, and data from forty-two (42) semi-structured interviews with individuals working for organisations with experience in research data sharing.

This data will be relied upon not only to make arguments about the current regulatory dynamics of this particular space, but also to present the historical development of research data sharing in the UK public sector over the last four decades (1980-2020).[103] For example, the role of particular organisations like the Department for Work and Pensions (DWP) at different periods of developing research data sharing as an activity will be examined, by reference to broader political narratives and developments.[104] Delving into the historical background of a space has been a common approach in the regulatory space tradition to understand the path-dependency of specific developments from the perspective of an 'evolutionary analysis'.[105] While various regulatory space studies do not provide such a historical background, the particularities of this space, i.e. the lack of sufficient information

[101] O Fioretos, T G Falleti and A Sheingate, 'Historical Institutionalism in Political Science' in theirs (eds), *The Oxford Handbook of Historical Institutionalism* (OUP 2016) 3.

[102] J March and J Olsen, 'The New Institutionalism: Organizational Factors in Political Life' (1984) 78(3) The American Political Science Review 734, 737.

[103] On the acknowledgment of the 'historical evolution and setting of regulation' by regulatory space approaches see M Lodge, 'Regulation, the Regulatory State and European Politics' (2008) 31(1) West European Politics 280, 290. Chapter 4 elaborates on these issues.

[104] *Infra* chapter 4, A-C.

[105] M Thatcher and D Coen, 'Reshaping European regulatory space: An evolutionary analysis' (2008) 31 West European Politics 806.

about research data sharing decisions,[106] make such an approach fruitful. This analysis will shed light on the way in which regulatory issues have been defined and specific organisational actors have succeeded in demarcating the boundaries of the space. With that in mind, I will be offering here a preliminary identification of the actors populating the research data sharing space.

Before proceeding with identifying the different types of organisational actors, it should be acknowledged that this task comes with certain limitations. A comprehensive regulatory space analysis would require 'systematic identification and examination of *all* relevant players (…) demanding an extremely thorough knowledge of the field'.[107] The activity of administrative data sharing for research purposes, however, as mentioned previously, is a relatively recent one in the UK and knowledge about its operation is far from comprehensive. Hence, the aim of this work is to make claims about the way in which this space is shaped by the *key* actors populating it, as well as the *main* forces that inform the exercise of regulatory power through the relationships between the actors.[108] This shall be treated as a high-level overview of the 'playing field' that will be further complemented and illuminated in chapter 4 by reference to the participation of specific organisational actors to the historical development of research data sharing in the UK.

Interestingly, the main actors in this space do not seem to fall neatly within the general, and often assumed as indisputable, categories of *regulators* and *regulatees*. Canning and O'Dwyer, for example, draw a distinction between 'pressurising constituents' (regulators) and 'targeted actors' (regulatees).[109] Chowdury and Wessel provide an even more specific categorisation by classifying actors into 'principals', 'participants' and 'residual actors'.[110] 'Principals' formulate the rules, 'participants' receive and implement or enforce them and 'residual actors' only follow them. This is neither to say that a general distinction between regulators and regulatees does not make sense, nor that the existence of hierarchical relationships within the

[106] *Supra* A in this chapter.

[107] Kaye and Gibbons (n 98) 113.

[108] Following the approach by Kaye and Gibbons *ibid* in mapping the regulatory space for genetic databases and biobanks in England and Wales.

[109] Canning and O'Dwyer (n 49) 170.

[110] Chowdury and Wessel (n 86) 339.

space is to be precluded.[111] Nonetheless, actors involved in research data sharing seemingly conform to the regulatory space assertion that actors constantly shift roles,[112] cautioning against a rigid and static role classification.

To give but a few examples, data custodians seem to be assuming different roles, at different times, and under different circumstances. On the one hand, they are 'receivers' of legal rules, empowering them to share or restricting them from sharing administrative data. On the other hand, they act as 'principals', formulating themselves organisational rules and practices which inform individual data sharing decision-making. Researchers may also assume different roles. Being members of organised groups, e.g. universities or research centres, they may operate in practice as 'pressurising constituents' by aiming to leverage their expertise and organisational alliances to steer decisions in their favour. At the same time, they are rule-followers or 'targeted actors' when it comes e.g. to demonstrating compliance with privacy safeguards or other requirements within a data access negotiation with a data custodian. By allowing challenges to the conventional divide between 'regulators' and 'regulatees', a regulatory space approach bears promise to illuminate the complex relationships between the actors in the research data sharing space.

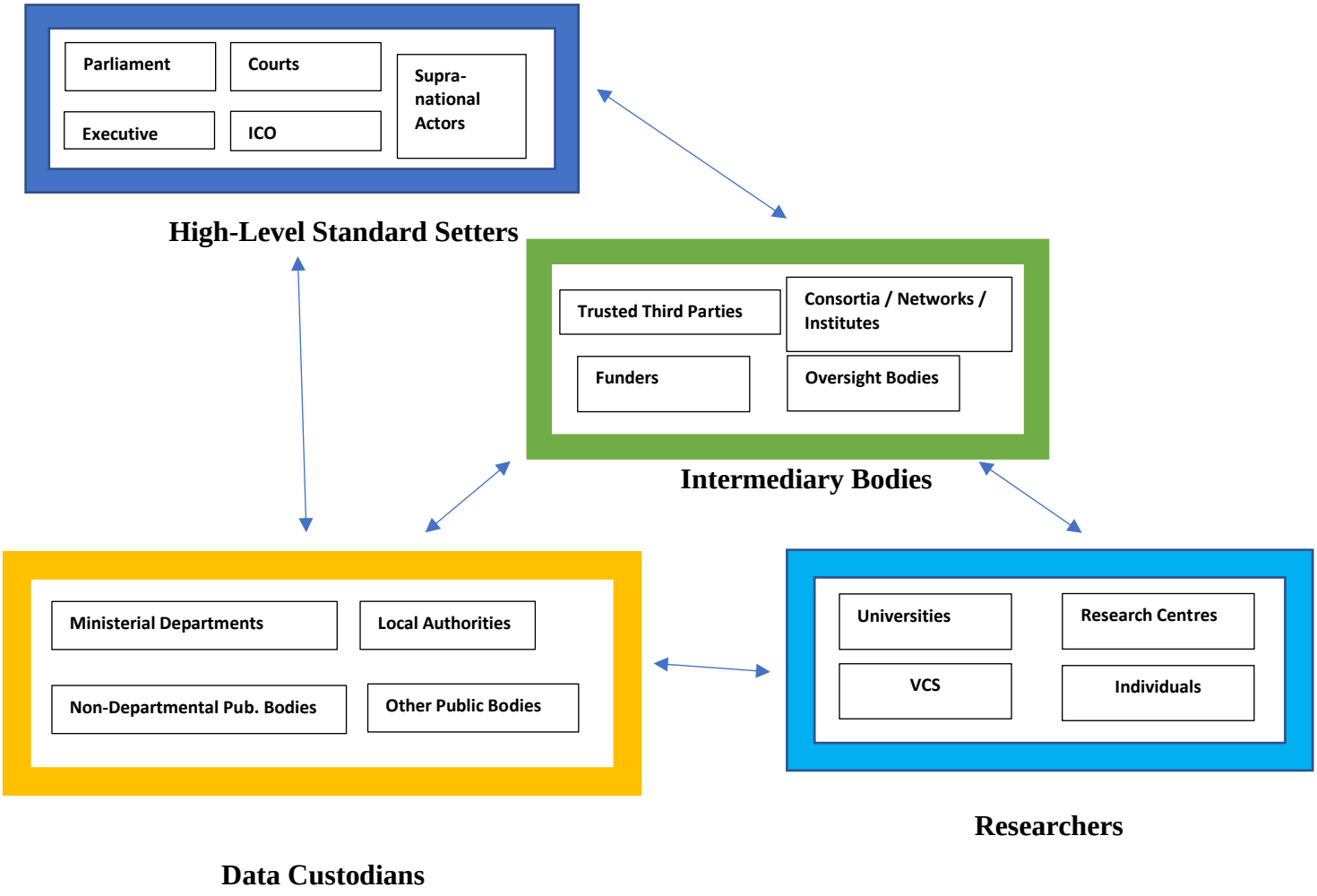
Arguably, the occupation of different roles by the actors is not only contingent on different 'times' or 'stages' of regulation, but also on the different regulatory resources to which the actors enjoy access. With that in mind, this thesis will not pre-impose a conceptualisation of different roles of actors, but shall aim to describe them *thematically* by reference to their organisational identities as *high-level standard-setters*, *data custodians*, *researchers* and *intermediary bodies*. This choice is consistent with scholarly accounts that have demonstrated the multiplicity of the role of e.g. the State as both a rule-maker and a rule-taker within such fields as transnational private regulation.[113] It is also a typology that proceeds from the formal capacities of the actors, since it is likely that some functional overlaps are likely to happen in practice. For example, the UK Statistics Authority is an oversight body when it comes to

[111] *Cf.* Chowdury and Wessel, *ibid* 348 on the relationships between actors in the space being 'necessarily non-hierarchical in nature (…) because the actors are not self-identified members operating within a well-defined and well-ordered regulatory system, which is based on legal rules'.

[112] Hancher and Moran (n 47) 152: 'The allocation of roles between rule makers, enforcers, and bearers of sectional interests constantly shifts, again obeying no obvious public-private dichotomy'.

[113] J Braithwaite and P Drahos, *Global Business Regulation* (CUP 2000).

determining compliance of research data sharing partners with legal requirements, but also functions as a high-level standard setter in its capacity as the creator of a 'code of practice' that governs research data sharing for public bodies in the UK.[114] The typology should also not be read too broadly; e.g. all bodies are 'intermediaries' when they mediate the communication between two other parties, yet 'intermediaries' here specifically refers to actors with a specific organisational mission to facilitate and support the activity of research data sharing. The following diagram depicts the main actors populating the research data sharing regulatory space:



This diagram indicates the complex, two-way relationships of influence between the parties. 'High-level standard-setters' encompass here both the three core domestic branches of state authority and supranational bodies. The category is defined here broadly to encompass both judicial bodies, ministerial departments, the Parliament and governmental agencies; the

[114] UK Statistics Authority (n 72).

common denominator between these bodies is that they are public sector bodies or bodies of a public character that are tasked with producing, monitoring or enforcing standards which govern administrative data sharing for research purposes. Examples of high-level standard-setters may include ministerial departments such as the Department for Business, Energy and Industrial Strategy (BEIS) in the UK or the Information Commissioner's Office. These actors set standards to be followed by data custodians but are also informed by the latter's organisational practices in doing so.

Through the regulatory space lens, this thesis will be interrogating the interactions between these actors in the process of developing research data sharing institutions without neglecting potential functional overlaps among their actual organisational activity. For example, 'data custodians' are the public-sector bodies responsible for handling access to administrative datasets for various, including research, purposes. It is important, though, to note a potential overlap between 'data custodians' and 'high-level standard-setters'. Some public bodies may wear both 'hats': on the one hand, they may be producing codes of conduct that are applicable to public-sector data sharing, while, on the other, they may be subject to pressures from the broader political environment about their own data sharing decisions. The ONS in England is a relevant example, substantially influencing the formation of cross-governmental data sharing norms during the drafting of the DEA 2017, while operating as a data custodian for various datasets currently utilised for administrative data research.[115] Nonetheless, the majority of data custodians, ranging from ministerial departments to local authorities, and the NHS, do not operate as standard-setters and are only considered here in their capacity to authorise or reject administrative data disclosures. 'Researchers' aim to influence data custodians by leveraging their expertise to secure a favourable data access decision, while they are also subject to the requirements imposed by custodians within this negotiation. Finally, 'intermediary bodies' mediate the relationship between custodians, researchers and high-level standard-setters by performing technical tasks (e.g. data matching or data linking)[116] or producing policy guidance

[115] Government Statistical Service, 'Introducing the Administrative Data Methods Research Programme' (1 July 2019) <https://gss.civilservice.gov.uk/news/introducing-the-administrative-data-methods-research-programme/>.

[116] H Morris et al, 'Challenges of administrative data linkages: experiences of Administrative Data Research Centre for England (ADRC-E) researchers' (2018) 3(2) International Journal of Population Data Science 97.

(e.g. reports recommending reforms to government data sharing policies)[117] with a view to enabling wider access to administrative data for research purposes. This is a loosely constructed category of actors including funding bodies such as the Wellcome Trust and the UKRI, research networks such as the ADRN, oversight bodies such as the UK Statistics Authority and bodies with technical expertise in research data sharing such as the NHS Wales Informatics Service. The following sub-sections elaborate on the above actors and their roles in the research data sharing space.

i. High-level standard-setters

Like most social activities, research data sharing is subject to the regulatory activity of high-level standard-setters, be that through the setting and enforcement of standards or other types of attempted behaviour modification. A meaningful initial distinction can be drawn here between *domestic* and *supranational* actors.

With *domestic* actors, relevant capacities are exercised by both the core branches of government and the independent public authority that is responsible for overseeing the protection of personal information in the UK, i.e. the Information Commissioner's Office (the ICO). As discussed in the introduction, the UK Parliament has legislated on a number of occasions to prescribe the conditions under which administrative data can be shared for research purposes, with the DEA 2017 and the Data Protection Act (DPA) 2018 being the most recent Acts. In respect of the judiciary, there is generally an abundance of data protection and privacy case-law on public-sector data sharing. The Courts have been willing to strike down legislation that is not compatible with the protection of fundamental data subject rights.[118] In the area of *research* data sharing, however, the role of the Courts is relatively limited so far, with no precedents addressing the legality of this particular conduct. On the side of the executive branch of the Government, BEIS has been the main actor aiming to cultivate a research data sharing culture in the UK public sector. BEIS has been pursuing this either

[117] REFORM, 'Sharing the Benefits: How to Use Data Effectively in The Public Sector' <https://reform.uk/research/sharing-benefits-how-use-data-effectively-public-sector>.

[118] E.g. *Christian Institute v Lord Advocate* (2016) UKSC 51 where the Supreme Court ruled that the data sharing powers in Part 4 of the Children and Young Persons (Scotland) Act 2014 were incompatible with Article 8 of the European Convention on Human Rights.

through its policy agendas or its multi-million investments in such data-intensive research-facilitating networks as the ADRN.[119] Finally, yet importantly, the ICO is the only regulatory body with enforcement and sanctioning powers when data sharing by public authorities breaches data protection legislation. As observed by Kaye and Gibbons with regard to medical research,[120] the ICO have, similarly, not exercised their powers extensively with regard to administrative data sharing for research. The ICO prefers instead to issue general data sharing guidance and rely on the voluntary compliance of data custodians.[121]

In the case of such *supranational* actors as the EU or Council of Europe institutions, their involvement in this space comprises mainly of the formation of formal legal standards that are applicable to the research data sharing context. For example, negotiations between conflicting interest groups within the European Parliament determined the final text of the EU General Data Protection Regulation and its impact on social science research.[122] In the same sense, the case-law of the European Court of Human Rights significantly informs the interpretation of data protection laws by data custodians and domestic Courts.[123]

ii. Data custodians

Data custodians are central actors in the research data sharing space. As they are processing administrative datasets to perform functions in the public interest, they are entitled to decide whether to allow access or not for research purposes. There are numerous types of data custodians in the UK public sector, considering that the latter is comprised of 25 ministerial departments and more than 400 other public bodies, i.e. executive agencies, non-departmental public bodies and non-ministerial departments.[124] If one adds local authorities and health and social care bodies to this list, it is clear that a comprehensive examination of custodians as actors becomes a demanding task. As previously mentioned, there is also a lack of concrete

[119] *Supra* chapter 1, A.

[120] Kaye and Gibbons (n 98) 120.

[121] ICO (n 10).

[122] L Stevens, 'The Proposed Data Protection Regulation and its Potential Impact on Social Sciences Research in the UK' (2015) 2 European Data Protection Law Review 97.

[123] J Bell et al, 'Balancing Data Subjects' Rights and Public Interest Research: Examining the Interplay between UK law, EU human rights law and the GDPR' (2019) 1 European Data Protection Law Review 43, 46.

[124] Cabinet Office, 'Departments, agencies and public bodies' <https://www.gov.uk/government/organisations>.

evidence in respect of the number of research data sharing requests that public bodies receive, as well as their decisions in response to them.[125] Hence, the present study will, inevitably, be selective in its examination of data custodians. Some prevalent examples include large ministerial departments like the Department for Work and Pensions (DWP) or the Ministry of Justice (MoJ),[126] as well as smaller but experienced in research data sharing public bodies like the ONS or the NHS Wales Informatics Service.[127] In respect of the selection of particular data custodians in this study, as chapter 3 will explain, the analytical and methodological approach of this work did not seek to produce a statistically representative sample of custodians or other actors in the space, but a 'theoretically meaningful one'.[128] Chapter 3 elaborates on the construction of this study's sampling strategy, the specific organisations that are part of my sample and my criteria for inclusion in that regard.

iii. Researchers

Researchers are another group of actors that is central in this space. Either via an intermediary body or directly, they interact with data custodians and seek to achieve a favourable data sharing decision. There are, arguably, many cases of individual researchers who have negotiated access to administrative data, with only a few cases of both success and lack of it having come to the attention of this author.[129] Nonetheless, in most cases, however, the term

[125] A recent effort to map the existing data sources for administrative justice research in the UK suggests further investigation into the dynamics of data ownership and acquisition by researchers, K Chatsiou et al, *Data Sources for Administrative Justice Research: A Preliminary Scoping Report* (2017) <https://doi.org/10.6084/m9.figshare.5303317.v2>.

[126] On the ADRN's struggles to access DWP data see Research Project Management (n 31) 3, 7; on the MoJ's practice in establishing a datalab to enable research that evaluates policy interventions to reduce re-offending see F Lyon et al, 'Opening access to administrative data for evaluating public services: The case of the Justice Data Lab' (2015) 21(2) Evaluation 232-47.

[127] National Institute for Social Care and Health Research Wales (NISCHR) (n 23) 3.13: 'Maximising the use of routine data can play a key role in enhancing the evidence base and offering innovative approaches to policy development and evaluation'.

[128] *Infra*, chapter 3, B-C.

[129] C Gill, *Administrative justice and the control of bureaucratic decision-making* (PhD thesis, University of Glasgow 2016) 6.34: 'the Department for Work and Pensions declined the researcher's request for access'; U Deetjen, *Internet use and health: a mixed methods analysis using spatial microsimulation and interviews* (DPhil

will refer to organised and well-established research groups such as the Population Data Science Group at the Swansea University or the Scottish Centre for Administrative Data Research at the University of Edinburgh.[130]

The particular interest of researchers in accessing the data is contingent on their organisational identity. With academic researchers, the aim will often be the development and support of an independent research agenda that will involve data-driven analysis to achieve scientific breakthroughs. For instance, academics within the Health Informatics Research Unit (HIRU) at the SAIL Databank benefited from access to administrative datasets in developing 'population data science' as a distinct field of enquiry within data science and informatics.[131] Other organisations, e.g. voluntary and community sector organisations (VCS) or social enterprises, might be more interested in accessing administrative data for statistical analysis and performance evaluation purposes. A relevant example comes from the Justice Data Lab (JDL) in the UK Ministry of Justice. The JDL accepts applications by VCS organisations willing to 'measure the impacts of their offender rehabilitation effort'.[132] Researchers negotiate with data custodians on various issues. The issues range from the construction of their research questions and their relevance to the custodian's policy functions, to the number of data records that will be disclosed and the appropriate safeguards to ensure lawfulness and social acceptability of research.[133]

iv. Intermediary Bodies

Finally, yet importantly, a wide range of other organisations mediate relationships between the previous actors in the interest of facilitating research data sharing. This category of

thesis, University of Oxford 2016) 339 on an 18-month long negotiation between the researcher and the Health and Social Care Information Centre (now NHS Digital).

[130] Population Data Science, 'Pioneering Population Data Science' <https://popdatasci.swan.ac.uk/about-us/>; Scottish Centre for Administrative Data Research, 'Data for social good' <https://www.scadr.ac.uk/administrative-data/data-social-good>.

[131] K McGrail and K Jones, 'Population Data Science: The science of data about people' (2018) 3(4) International Journal of Data Population Science.

[132] Lyon et al (n 126) 235.

[133] J Teng et al, 'Sharing linked data sets for research: results from a deliberative public engagement event in British Columbia, Canada' (2019) 4(1) International Journal of Data Population Science.

intermediary bodies comprises research funders, research networks or institutes, independent liaisons known as 'trusted third parties' and oversight bodies. While intermediaries come in different forms, e.g. as public bodies or not-for-profit think-tanks, they share an organisational mission to facilitate administrative data sharing for research in the 'public interest'.[134] They do that by either enhancing communication between the other participants of the space or influencing broader political and policy developments through promoting their own agendas about data-intensive research. Intermediaries may use a variety of tools in doing that, from grant calls to form administrative data research consortia,[135] to public campaigns about the benefits from research data sharing for the broader public[136] and policy reports that propose research-data-sharing-related reforms in the public sector.[137]

Research funders in the area of social sciences, e.g. the Economic and Social Research Council (ESRC), the Nuffield Foundation and the Wellcome Trust are interested in enabling research data sharing synergies between data custodians and academic researchers. Their interest stems from their mandate to stimulate the production of high-quality, robust social-scientific research with a potentially major impact on real-world, often policy-related problems. Funders may employ different strategies in promoting their aims. They might commission research reports on the benefits of open data research or the appropriate design of a data access policy model.[138] A major strategy involves the creation of a distinct research consortium or initiative with the explicit aim of promoting research data sharing. For example, the UK Administrative Justice Institute (UKAJI) was established by the Nuffield Foundation in 2014 to 'kickstart the expansion of empirical research on administrative justice in the UK'.[139] Within the particular context of evaluating administrative justice operations and decision-

[134] NISCHR (n 23) 3.13.

[135] ADR UK, 'Administrative Data Research UK' <https://www.adruk.org/>.

[136] Understanding Patient Data, 'Putting people at the centre of decisions about patient data' <https://understandingpatientdata.org.uk/>.

[137] REFORM (n 117).

[138] Wellcome Trust, 'Data Policy Models for Funding Bodies' (September 2018) <https://wellcome.ac.uk/sites/default/files/data-policy-models-for-funding-bodies-eagda.pdf>; Expert Advisory Group on Data Access, 'Governance of Data Access' (June 2015) <https://wellcome.ac.uk/sites/default/files/governance-of-data-access-eagda-jun15.pdf>.

[139] UK Administrative Justice Institute, 'About UKAJI' <https://ukaji.org/about/>.

making, the UKAJI aspires to 'link policy, practice and research communities'. Widening access to administrative data for researchers is a major aim for the UKAJI in that regard.[140]

Beyond funders and research networks or institutes, there are intermediary bodies which are tasked with providing assurances in respect of the operational realities of data sharing. The main example in this area is the *trusted third party* (TTP) system. A TTP is an 'independent organisation that acts as a liaison between two or more collaborating' organisations,[141] being responsible for the promotion of mutual trust between data sharing collaborators and preventing re-identification of data subjects throughout the research process. Bodies with relevant organisational capacity and information governance expertise assume the role of a TTP, e.g. the ONS in England, the NHS Wales Informatics Service in Wales or the Northern Ireland Statistics and Research Agency (NISRA) in Northern Ireland. TTPs perform a variety of technical tasks to facilitate research data sharing, such as de-identification and linkage,[142] i.e. the removal of identifying characteristics in the data and the linking of a dataset with other datasets that will allow both illuminating statistical findings and preserve confidentiality of data subjects. Lastly, there is also a body with a legal mandate to oversee the conformity of both data custodians and researchers with such broad legal principles as transparency, proportionality and confidentiality. This is the UK Statistics Authority, which is required under section 70 of the DEA 2017 to enforce the adherence of all parties to the principles contained in its Code of Practice.[143] While we are still in the first stages of this body's operation, similar bodies in the health data sharing space have been instrumental in allaying data custodians' worries about confidentiality risks.[144] Having outlined the key actors populating the research data sharing space, the next section introduces the concept of 'research data sharing institutions'.

[140] Chatsiou et al (n 125).

[141] V den Braak et al, 'Trusted third parties for secure and privacy-preserving data integration and sharing in the public sector' (2012) Proceedings of the 13th Annual International Conference on Digital Government Research 135, 138.

[142] K Harron et al, 'Challenges in administrative data linkage for research' (2017) 4(2) Big Data & Society.

[143] UK Statistics Authority (n 72).

[144] M J Taylor, 'Legal Bases for Disclosing Confidential Patient Information for Public Health: Distinguishing between Health Protection and Health Improvement' (2015) 23(3) Medical Law Review 348.

C. Institutionalisation in the Research Data Sharing Space

Administrative data sharing for research purposes is a heavily routinised and technical activity, characterised more by regularised everyday processes than 'momentous events'.[145] Data access requests have to be reviewed thoroughly by various decision-makers or intermediaries to ensure conformity with legal and technical requirements, as well as contribute to the 'public interest' as defined by a data custodian.[146] The regulatory space lens adopted in this study sheds light on the role of such various actors in developing, interpreting or displacing these legal and technical requirements that govern research data sharing in the UK.

Neo-institutional research has demonstrated that civil servants routinely follow in their decision-making certain *institutions,*[147] defined as an assemblage of 'rules, procedures, organisational standards, governance structures, conventions and customs',[148] as well as 'symbol systems, cognitive scripts and moral templates'.[149] While some socio-legal studies think of legal rules as a paradigmatic example of institutions, one could use the example of an organisational 'research data sharing policy' as a prime example of a research data sharing institution.[150] Institutions have *cognitive, regulative* and *normative* dimensions.[151] They reflect 'taken-for-granted' logics behind particular social activities,[152] embody disciplinary mechanisms that secure compliance and aid individual decision-makers to understand what is the appropriate response to a given situation based on their position and responsibilities.[153] To refer to the example of a research data sharing policy, this would involve definitions of the

[145] Powell and Colyvas (n 66) 285.

[146] NISCHR (n 23) 3.13.

[147] Powell and Dimaggio (n 59) 147; J Meyer and B Rowan, 'Institutionalized organizations: formal structure as myth and ceremony' (1977) 83 American Journal of Sociology 340-363; J Meyer and W Scott, *Organizational Environments: Ritual and Rationality* (Sage 1983).

[148] Powell and Dimaggio (n 59) 147.

[149] Hall and Taylor (n 60) 940.

[150] E.g. see the relevant policy developed by the ONS in England in Office for National Statistics, 'ONS research and data access policy' <https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy>.

[151] R Scott (n 67) 33.

[152] Jonnergård and Larsson-Olaison (n 49) 464.

[153] March and Olsen (n 102) 738.

types of data processing that an organisation would consider as compatible with the 'public interest', as well as provisions about the oversight of data sharing agreements between custodians and researchers and the penalties (e.g. interruption of data access) in case of non-conformity.[154] By associating 'stable, valued, recurring' types of behaviour with purposes that transcend individual intentions,[155] institutions render particular decision-making considerations as relevant or irrelevant.[156] As Edelman argues, organisations incorporate 'institutionalised models' for both rational, cost-benefit, reasons and because 'certain actions, forms, or rituals come to be understood as proper and natural'.[157] A research data sharing policy, for example, could proceed from the premise that research data sharing needs to be directly relevant to the functions of the data custodian before it can be considered for approval.[158]

Existing knowledge about cautious decision-making in the UK public sector provides reasons for believing that data sharing decision-makers are no exception to the influence of institutions.[159] Disagreements, as outlined in the introduction,[160] about the role of law or organisational culture in fuelling a 'culture of caution' in access decisions imply the existence of cognitive and normative frameworks that frame the activity of data sharing in particular ways, 'guiding, constraining and empowering' the behaviour of data sharing participants.[161] There seems to be a variety of data sharing institutions at play e.g. legal rules prescribing data sharing safeguards, organizational procedures structuring custodian decision-making,

[154] UK Statistics Authority (n 72).

[155] S Huntington, *Political Order in Changing Societies* (Yale University Press 1988) 12.

[156] March and Olsen (n 102) 738; L Edelman, 'Legal Endogeneity and the Limits of Equal Opportunity' (2004) <https://memphis.edu/benhooks/creative-works/pdfs/edelman.pdf> 6.

[157] Edelman *ibid* 6.

[158] HM Revenue and Customs, 'Research at HMRC' <https://www.gov.uk/government/organisations/hm-revenue-customs/about/research>: 'Proposals are then evaluated against these criteria: contribution to HMRC functions, long-term benefit to HMRC, research design and likely impact of the research, feasibility of the research'.

[159] Law Commission (n 2); Research Project Management (n 31); Thomas and Walport (n 2); Laurie and Stevens (n 7); ADRC Scotland, 'From a Culture of Caution to a Culture of Confidence: Lessons Learned From Implementing the Public Records (Scotland) Act 2011' <https://www.research.ed.ac.uk/portal/files/27327583/Publicity_Events_ADRC_Workshop_FINAL.pdf>.

[160] *Supra* chapter 1, 15-17.

[161] R Scott (n 67) 8.

economic rationalities of prioritising resource allocation within custodian bodies or political ideas about the accountability of data custodians.[162] Such institutions are reflected in and transported by multiple 'carriers' such as codified instruments (e.g. statutory acts like the DEA 2017), routines (e.g. routine assessment of the contribution of a request to the public interest within the civil service) and cultures (e.g. ideas and assumptions about public sector data ownership).[163] Hence, it is important to clarify one's analytical focus.

Even if the primacy of law as an influence on organisational decision-making has often been challenged in socio-legal studies,[164] regulation studies often focus on the question of compliance with *legal* institutions.[165] More specifically, there is a tendency of regulatory space studies to highlight the significance of legal or regulatory requirements, which are formally articulated and codified, in their explanatory accounts of power relationships within a particular space.[166] To draw on a prominent example, Colin Scott explains how private entities dominate 'the interpretation of regulatory obligations'.[167] Similarly, Canning and O'Dwyer, being interested in the re-negotiation of the regulation of Irish professional accounting, aim to 'unveil how potentially powerful the accounting profession can be in determining both the content of regulations and the means through which they may be enforced'.[168] In doing so, they take the 'practical meaning of legislation' to be a key issue of negotiation.[169] Jonnergård and Larsson-Olaison also study the reshaping of corporate governance regulation in Sweden by reference

[162] Hancher and Moran (n 47) 156; C Scott (n 47) 289.

[163] R Scott (n 67) 52: 'Institutions, whether regulative, normative, or cognitive perspectives are stressed, are embedded in various types of repositories or carriers'.

[164] J Black, 'New institutionalism and naturalism in socio-legal analysis: Institutionalist approaches to regulatory decision making' (1997) 19(1) Law & Policy 51, 53 on law as 'not determinative' within organisational decision-making; M Hertogh, *Nobody's Law: Legal Consciousness and Legal Alienation in Everyday Life* (Palgrave 2018) 131: such key Dutch public law principles as 'legality' and 'equality' are of low relevance to the actual decision-making of administrative officials in the Indonesian Quarter in Zwolle.

[165] E.g. D McBarnet and C Whelan, 'The Elusive Spirit of the Law: Formalism and the Struggle for Legal Control' (1991) 54(6) Modern Law Review 848.

[166] Reflecting the neo-institutionalist trend to operate at the macrolevel, focusing on the role of an institutionalised environment ('the state' or 'law' or 'constitutional orders') in legitimising organisations and their structures; see L Zucker, 'Institutional Theories of Organization' (1987) 13 Annual Review of Sociology 443-464.

[167] C Scott (n 47) 300.

[168] Canning and O'Dwyer (n 49) 169.

[169] *Ibid.*

to the reform of domestic corporate law. They elaborate on the ways in which the 'international financial market discourse exerts substantial influence on the referral process and development of the new corporate governance code'.[170] In all these examples, analytical focus is placed on the formation and interpretation of *codified* and *formal* legal or regulatory instruments. Is such an emphasis on codified and formally articulated institutions unjustified or problematic?

There are, certainly, reasons behind focusing on this type of institutionalisation, i.e. the formation, transmission and prevalence of institutions as reflected in formal and codified regulatory requirements.[171] One of the central interests of regulatory space studies is to assess how power is exercised within the interrelationships of various actors in a particular space. For social-scientific purposes, however, 'power' is not observable; there is a need to find appropriate proxies, which can be understood empirically, that will allow us to assess who wields the relevant kind of power in the given situation. One helpful way of making such assessments is, indeed, through studying the processes through which actors are 'influencing and shaping substantive and procedural rules that govern regulatory outcomes within the specific regulatory space'.[172] To give an example from the research data sharing context, the English Office for National Statistics (ONS), despite being administratively and financially 'smaller' than many other government departments, managed to mobilise its scientific expertise to effectively influence the contents of the Digital Economy Act (DEA) 2017 which establishes cross-governmental research data sharing norms in the UK public sector.[173] There is, admittedly, significant potential in reflecting upon examples like this to understand how research data sharing institutions have been shaped in this space. Is there anything wrong, then, with focusing on formally articulated and codified regulatory requirements?

This thesis answers this question in the affirmative, arguing that the regulatory space literature tendency to confine its analysis of regulatory power to the capacity to shape the

[170] Jonnergård and Larsson-Olaison (n 49) 460.

[171] P Tolbert and L Zucker, 'What are Microfoundations? Why and how to study them?' in P Haack, J Sieweke and L Wessel, *Microfoundations of Institutions* (Emerald 2020) 4.

[172] See Chowdury and Wessel (n 86) 339.

[173] Sections 64 – 70 of the DEA empower all public authorities to disclose data they hold 'in connection with their functions' for research purposes, largely adopting the framework developed by the ONS within its discretion to share data from the 'approved researcher' gateway under section 39(5) of the Statistics and Registration Act (SRSA) 2007.

interpretation of formally articulated obligations and rules is overly restrictive.[174] As Scott himself has explained, understanding 'who can enter' a specific regulatory space and 'what is the appropriate scope of the regulatory issues' requires insight into a broader set of institutions such as: 'the cultural environment within which (organisations) work, their standard operating procedures (…) (and) the customary assumptions which govern their interactions (…)'.[175] Hancher and Moran similarly conceive of a broader set of 'regulatory institutions' that actors may create or copy, drawing on more dynamic, cultural or customary assumptions which govern their interrelationships.[176] Such a broader definition of institutions is also adopted in this thesis while examining how interactions between organisational actors delimit the scope of this specific regulatory space. In shaping such institutions, actors in the research data sharing space draw on a set of key resources, as suggested by regulatory space analysis, i.e. information, organisational capacity, wealth and legal authority.[177]

Some data sharing institutions may be *settled, durable* and *articulated*: legal rules, organisational procedures or formal networks between organisations with a view to facilitate research data sharing. Other institutions may be *non-articulated, fluid* and *dynamic:* different subjective interpretations of legal requirements, customary assumptions about the value of social research, conventions of approving particular types of research requests or cognitive processes of knowledge exchange between data analysts working for different actors. Edelman draws a similar distinction between 'structures' and 'consciousness' in explaining the function of the 'legal environment of organizations'.[178] She associates 'rules, practices or procedures' with 'structures' and 'beliefs or ideas about technology, governance, management efficiency and (…) understandings about the meaning of law and ideas about how to make use of legal rules and rights' with 'culture'.[179] As Scott notes, such 'institutions (…) as culture and

[174] Even if the resources include other forms of authority such as 'information, wealth and organisational capacities', see C Scott (n 47) 284. On this point, also see Dove (n 12) 83 on the need to distinguish between a legal mandate and 'actual regulatory authority' in the context of health research regulation.

[175] *Ibid.*

[176] Hancher and Moran (n 47) 156, 159.

[177] Scott (n 47) 287.

[178] L Edelman, *Working Law: Courts, Corporations and Symbolic Civil Rights* (Chicago University Press 2016) 155.

[179] *Ibid* 163.

organisational form'[180] delimiting the regulatory issues in a space are to be understood empirically.[181] It is suggested in this thesis that the more dynamic institutions are not less important or influential than the articulated and durable ones. The former merit analytical focus as much as the latter.

To do so, the present study will identify both 'legal' and 'non-legal' institutions that filter access to administrative data for social-science research purposes.[182] The historical development of such institutions in the UK research data sharing space will be placed in its legal, social, political and economic context.[183] Apart from a heavily routinised and institutionalised activity, however, research data sharing is also an activity that involves multiple interactions and communications between organisations, mainly between data custodians, requesting researchers and intermediary bodies.[184] The regulatory space lens, departing from fixed conceptions of power balance between 'regulators' and 'regulatees', facilitates an empirical understanding of both high-level and routine, commonplace interactions between organisational actors which populate a particular space. The following and final section elaborates on the interactions between various actors, with a particular emphasis on micro-level and routine interactions between data sharing participants, as a mechanism through which decisions about data access are governed in the research data sharing space.

D. Regulatory Interactions in the Research Data Sharing Space: From Macro to Micro

Interactions involving relationships of 'conflict' and 'bargain' between actors are at the heart of regulatory space studies which seek to illuminate the construction of a particular space. While organisational actors function simultaneously as 'actors and as arenas for interaction by

[180] C Scott (n 47) 288.

[181] *Ibid.*

[182] Black (n 164) 60.

[183] On the acknowledgment of the 'historical evolution and setting of regulation' by regulatory space approaches see M Lodge (n 103) 290. Chapter 4 elaborates on these issues.

[184] As it will be shown in chapter 4, high-level standard-setters, while setting the broader terms of data sharing in place, interact less often with the other actors in this space.

individuals',[185] regulatory space studies often prioritise high-level interactions between mostly large organisations or professional communities.[186] Indeed, such studies often focus on interactions between key actors, leading to or revolving around 'momentous events' in regulatory development.[187] For example, Jonnergard and Larsson explain how the 'interpretive community' of Swedish business 'insiders' was influential in the reform of the Swedish corporate governance code through particular 'regulatory conversations'.[188] Similarly, Malsch and Gendron aim to understand 'processes of overt conflicts over preferences, agenda setting and discursive construction' in the context of accounting regulation in Canada.[189] Their findings about a particular 'form of allegiance' between regulated firms and the bodies overseeing them were crucially informed by 'very concrete episodes of agenda setting and overt confrontation'.[190] This emphasis is in line with neo-institutionalist interests in institutionalisation as a macro-level state, highlighting the role of 'leaders' of 'champions' within organisational interactions in triggering broader culture shifts or legal reforms.[191]

This approach in the regulatory space literature is not satisfactory for the purposes of the present thesis. Research data sharing decisions are often made by individuals or small teams of individuals, often analysts working for government departments, with which requesting parties and intermediary bodies have to interact on an ad-hoc basis and according to the requirements of each data custodian.[192] The idiosyncrasies of this multi-party relationship, structured around the minutiae of an everyday technical process performed by small groups of individuals, might not be adequately captured by a high-level view of the space and the interrelationships of some

[185] S Wood et al, 'The Interactive Dynamics of Transnational Business Governance: A Challenge for Transnational Legal Theory' (2015) 6(2) Transnational Legal Theory 333, 341.

[186] Hancher and Moran (n 47) 271.

[187] Powell and Colyvas (n 66) 285.

[188] Jonnergård and Larsson-Olaison (n 47) 460.

[189] B Malsch and Y Gendron, 'Reining in Auditors: On the dynamics of power surrounding an ''innovation'' in the regulatory space' (2011) 36 Accounting, Organizations and Society 456.

[190] *Ibid* 474.

[191] Tolbert and Zucker (n 171) 7.

[192] E.g. for justice-related data, the HMCTS clarify on their website the precise steps a research request should follow, HMCTS, 'Access to courts and tribunals for academic researchers' <https://www.gov.uk/guidance/access-to-courts-and-tribunals-for-academic-researchers>.

of its key decision-makers or high-level constituents.[193] For example, when a data analyst managing a particular dataset on behalf of a local authority develops a set of criteria to assess the extent to which the application of statistical methods that use a certain number of variables from that dataset can produce valuable insights, she arguably contributes to a collective understanding and definition of the 'public interest' in research data sharing. A high-level view of the space might not allow a sufficient problematisation of the role of such individuals, often working at a lower and more technical level, as 'boundary-crossing norm entrepreneurs'.[194] This role shall not be underestimated in a space where policy decisions and enthusiasm for digital experimentation may be more difficult to implement on-the-ground than envisaged by high-level policy-makers.[195]

This is not to disregard the importance of seeing the bigger picture in the research data sharing space. A high-level understanding of the historical development of research data sharing in the UK public sector is certainly significant and chapter 4 will be drawing on documentary and interview data to identify key moments, leading actors and dominant political discourses operating in the context of such development.[196] Nonetheless, considering the previously mentioned idiosyncrasies of research data sharing interactions, these macro lines of analysis (broader discourses, leading actors and key moments) will be complemented by a micro-level understanding of organisational interactions within the space. Since research data sharing institutions are created, interpreted, sustained and altered 'as they are enacted by individuals in concrete social situations',[197] emphasis will be shifted to the level of commonplace interactions between individuals negotiating data access, often involving representatives from two or more organisational actors. This can provide a rich understanding of how participants 'maintain or transform the institutional forces that guide daily practice'.[198]

---

[193] Regulatory space studies have been criticised for making more abstract and complex the task of understanding regulation, see J Black, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (2001) 54 Current Legal Problems 103, 109

[194] Wood et al (n 185) 341.

[195] M Veale, M van Kleek and R Binns, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' (2018) Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems 440.

[196] *Infra* chapter 4.

[197] Powell and Colyvas (n 66) 284.

[198] *Ibid.*

Regulatory space scholars have followed this approach in their studies. As Chowdury and Wessel stress, a focus on the 'micro-level dynamics' of specific aspects of a regulatory regime, 'enables us to identify certain regulatory trends and contextualise macro-level developments'.[199] In elaborating on how regulatory interactions between organisations 'define and redefine' a space,[200] Black called attention to 'micro-level conversations that may occur in formal or informal settings',[201] potentially extending beyond formally articulated standards reflected in legal and policy documents. Within such conversations, individuals not only adopt broader frameworks to their behaviour, but also contribute to the ongoing construction of these frameworks by introducing novel understandings and interpretations.[202] According to Black, such interactions are particularly important where discretion is a significant element of the regulatory process and when the task of regulation is 'uncertain'.[203] These criteria are met in the case of the research data sharing space in the UK public sector; data custodians have discretionary legal powers to disclose data to researchers,[204] whereas the complex technical character of data sharing, involving sophisticated statistical techniques of data cleaning and data matching, creates the need for agreement through 'intersubjective sharing of ideas and negotiations of meaning'.[205]

Hence, with relatively little knowledge being available about research data sharing conversations and interactions in the public domain, my analysis of empirical data will seek to unravel the process through which norms, cognitive frames, data science methods and political ideals emerge and structure data access decision-making.[206] Including both 'less powerful' members of organisations and more senior participants, the aim of the analysis will be to interrogate how are institutions shaped between the micro and the macro level, either emerging in commonplace interactions and then 'ratcheting upwards' or being 'pulled down' from

[199] Chowdury and Wessel (n 86) 347.

[200] Black (n 100) 188.

[201] *Ibid* 171.

[202] R Scott (n 67) 51.

[203] Black (n 100) 172.

[204] Department for Culture, Media & Sport (n 30).

[205] Black (n 100) 172.

[206] Powell and Colyvas (n 66) 285.

macro-orders like the legal framework and thereby structuring organisational and individual behaviour.[207]

Micro-level interactions in this space may involve various types of communication. For example, communication may foster facilitative relationships between trusted third parties and data custodians to ensure appropriate data security and confidentiality safeguards, as well as the broader social acceptability of administrative data research.[208] On the other hand, conflict and bargaining will often arise with regard e.g. to a request's aspiration to access data that data custodians would not like to see used to criticise their policy-making.[209] Similarly, tensions may emerge around the status of administrative data ownership; views expressed by high-level standard-setters or funders with regard to the status of such data as a 'public good' or 'public infrastructure'[210] are less likely to be shared by data custodians. In both cases, the space's constituents aim to leverage their organisational resources within their interactions with other data sharing participants in the interest of shaping norms, ideas, conventions and procedures.[211]

Considering the technical and complex activity of research data sharing, individuals with particular roles such as data analysts or information governance officials seem to be particularly important for the success of organisational efforts to (re)claim parts of the space. The role of professional and other expert groups in 'constructing the organisational field' and 'creating the categories, the norms, the rules and standards' has been highlighted before in the institutionalist literature.[212] Such groups present themselves as 'uniquely qualified' to define

[207] Tolbert and Zucker (n 171) 7.

[208] P Carter et al, 'The Social License for Research: Why Care.data Ran into Trouble' (2015) 41 Journal of Medical Ethics <https://jme.bmj.com/content/41/5/404>; Administrative Data Taskforce (n 25).

[209] Jones et al (n 31).

[210] House of Commons Science and Technology Committee, 'The big data dilemma: Fourth Report of Session 2015-16' <https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf> 20; P Arzberger et al, 'Promoting Access to Public Research Data for Scientific, Economic and Social Development' (2004) 3 Data Science Journal 135, 136.

[211] Several institutional scholars have highlighted the significance of communication for creating, shaping, and disrupting institutions, emphasizing that 'at its core, institutional theory is a theory of communication', see R Suddaby, 'How communication institutionalizes: A response to Lammers' (2011) 25(1) Management Communication Quarterly 183, 188.

[212] P DiMaggio, 'Constructing an Organizational Field as a Professional Project: U.S. Art Museums, 1920-1940' in W Powell and P DiMaggio (eds), *The New Institutionalism in Organizational Analysis* (Chicago University

regulatory problems,[213] aspiring to authoritatively provide solutions to such problems. In the research data sharing space, expertise in managing and analysing administrative datasets gives rise to the creation of communities of experts who collaborate across organisational boundaries and interpret authoritatively the 'rules of the game' within this space.

These communities may be formed on the basis of organisational identity, e.g. the representation of the interests of a public body, or on the basis of epistemic background, e.g. in statistical analysis and management of massive datasets. Based on these characteristics, individuals who participate in data access negotiations, either information asset owners working for data custodian bodies,[214] researchers or employees of intermediary organisations, seek to promote their frames of reference and defend what they see as appropriate and desirable. For example, an information asset owner representing a data custodian may invoke specialised terminology and sophisticated technical explanations to advocate for the appropriateness of a particular data sharing solution e.g. to argue for the introduction of additional restrictions to data access.[215] An administrative data researcher, interested in accessing a particular dataset, may provide examples of the risks emerging from non-using administrative health data to better understand the transmission risks of a disease within a community with a view to broadening the interpretation of 'public interest' as a condition for data access.[216] Through the interactions between individuals representing actors in the space, research data sharing institutions governing access decisions are enacted, practiced and, from time to time, fundamentally reconstructed.

Press 1991) 267; S Brint and J Karabel, 'Institutional Origins and Transformations: The Case of American Community Colleges' in Powell and DiMaggio (eds), *ibid* 337. Institutionalism is defined as the political science approach that focuses on the constraining and enabling effects of formal and informal rules on the behaviour of individuals and groups.

[213] R Scott (n 67) x.

[214] E.g. decision-makers working for the NHS Wales Informatics Service decide as trustees on behalf of specific health and social care data custodians in Wales, NISCHR (n 23) 8.

[215] 'Linking Administrative Data: recent experience within the Department for the Economy' (on file with the author) – note that the Department for the Economy (NI) has succeeded the Department for Employment and Learning (NI). This individual was approached for an interview, but preferred to provide the author with a (comprehensive) written statement of their view on the matter instead.

[216] K H Jones et al, 'The other side of the coin: Harm due to the non-use of health-related data' (2017) 97 International Journal of Medical Informatics 43.

In conclusion, this chapter has explained the appropriateness of the regulatory space concept as an analytical lens through which to understand the regulation of research data sharing decisions, proposing a further tripartite analysis of *actors*, *institutions* and *interactions*. It has been argued that this analytical framework sheds light on the idiosyncrasies of research data sharing in the UK public sector: multiple actors collaborate with one another over a long time, often repeating routinised technical processes performed on a lower decision-making level, while demonstrating conformity to organisational and legal requirements that are not necessarily formally articulated or codified. It will be for the following chapters of this thesis, i.e. chapters 4-7, to make empirically-informed arguments about the regulation of research data sharing decisions in the UK public sector through the presently discussed analytical lens. These arguments, through their reference to specific actors in their empirical context, will illuminate the understanding of the dynamics of the presently discussed regulatory space, facilitating an understanding of the 'rules of the game' that filter access to administrative data for research purposes in the UK. The next chapter presents the methodological framework from which this analysis stems.

3. STUDYING THE REGULATION OF RESEARCH DATA SHARING DECISIONS THROUGH AN IN-DEPTH, QUALITATIVE RESEARCH DESIGN

This chapter presents the core elements of the methodological framework adopted in the present thesis: an in-depth, qualitative research design, which employed fourty-two (42) semi-structured interviews with individuals working for organisations with experience in research data sharing and documentary data about the regulation of research data sharing decisions in the UK public sector. I begin by providing a micro-level 'anatomy' of research data sharing decision-making, fleshing out the key roles and job types involved in this activity, to clarify the focus of my empirical research. Then, I present my approach to data collection and analysis, while discussing how I aimed to enhance the 'validity' and 'reliability' of my findings, including through the design and deployment of a small-scale pilot survey. Finally, I also discuss limitations of my research design.

A. A Micro-Level 'Anatomy' of Research Data Sharing Decisions

As the previous chapter clarified, the present thesis does not seek to explain the *outcomes* of data disclosure decision-making, taken either individually or collectively.[217] The focus of this thesis, instead, lies on the *process* of decision-making, seeking to explain the way in which this activity has been shaped and governed over time in the United Kingdom. The previous chapter identified three dimensions to this inquiry: actors, institutions and interactions. Through their interactions, large organisational actors create, maintain and modify research data sharing institutions which structure and justify individual data sharing decisions from the perspective of their participants. Yet, the discussion so-far has not come close to the micro-level, i.e. the nuts and bolts of everyday data sharing decision-making: who are the individuals involved in decisions in practice? How do they communicate with each other? What are their organisational affiliations and professional qualifications? To inform discussion of my approach to data collection and demystify the intricacies of research data sharing decision-making, I begin with shedding some light on these questions. It is important to state from the outset that, considering the uncertainty about the number of data custodians, the requests they receive and their decision-making practices in the UK public sector, the presently offered 'anatomy' of data sharing decisions involves a certain degree of abstraction.

[217] *Supra* chapter 2, A.

'Data analysts' are the officials who are working most closely to the datasets. Although they may come from a variety of disciplinary backgrounds,[218] analysts share a common educational background which involves statistical knowledge and some familiarity with managing datasets.[219] Analysts are responsible for the core operational dimensions of research data sharing. These dimensions include such tasks as identifying the desired variables for a research project within the datasets, matching particular variables with other variables of interest in other datasets and 'cleaning' the datasets so that researchers can use them.[220] Beyond data custodians, data analysts also work for intermediary bodies. For example, the ADRC-NI employs 'Statistical Methodological Officers' (SMOs) who assist researchers in identifying the best datasets for their project's purposes and communicating their requests to the data custodians with accuracy and clarity.[221]

The second group, which also works relatively closely to the datasets, is that of 'Information Governance' (IG) officials. IG officials are responsible for ensuring that data sharing procedures are not only robust from an information security perspective, but also compliant with the law and, more broadly, with established standards of best practice.[222] Like data analysts, IG officials come from diverse backgrounds, which often include studies in law, systems architecture and information management. Since data custodians are responsible for the lawfulness and technical propriety of data sharing, IG officials are mostly working for data custodian bodies.[223]

---

[218] C Lindner, 'Where do Data Scientists Come From?' <https://medium.com/indeed-engineering/where-do-data-scientists-come-from-fc526023ace>.

[219] P Besse and B Laurent, 'From Statistician to Data Scientist' (2015) arXiv:1509.07804 [stat.OT]

[220] R Goerge and B J Lee, 'Matching and Cleaning Administrative Data' in M Ver Ploeg et al (eds), *Studies of Welfare Populations: Data Collection and Research Issues* (NAP 2002) 197.

[221] ADR UK, 'Administrative data is an invaluable resource for public good. Let's use it' (2019) <https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK_Annual_Report_2018-19.pdf>.

[222] T Keyser and C Dainty, *The Information Governance Toolkit: Data Protection, Caldicott, Confidentiality* (CRC 2016) 20.

[223] Although certain intermediary bodies also specialise in information governance, see D Ford et al, 'The SAIL Databank: building a national architecture for e-health research and evaluation' (2009) 9 BMC Health Services Research.

Finally, yet importantly, policy-makers are a significant group within the data sharing process, despite the fact that they are both physically distant and distant in terms of technical expertise and understanding from the datasets. While analysts advise about the feasibility of a proposed data disclosure and IG officials about the legal and technical safeguards that should be implemented before the disclosure takes place, policy-makers concern themselves with the overall desirability and appropriateness of the request.[224] Crucially, policy-makers assess whether the proposed disclosure is in the 'public interest', taking into account the potential benefits that the research could yield and the costs that the data custodian would have to incur to share the data. While public bodies can generally process administrative datasets for reasons not strictly confined to one of their functions that is in the public interest, e.g. in the context of a commercial transaction, the empirical chapters of this thesis will show that in the case of data sharing for research purposes it is deeply rooted in custodian cultural practice to justify their decisions by reference to the public interest.[225] Policy-makers dealing with this particular issue are only employed by data custodians, even if intermediary bodies also employ staff with responsibilities in policy-making and influencing the broader data sharing environment.[226]

Having outlined the key job types of officials involved in research data sharing decisions and their core responsibilities, the questions about the factors shaping such decisions become clearer: how do the different groups communicate with one another? How are officials performing a specific job within data custodian bodies influenced by officials performing the same job for intermediary bodies or researchers? What are the 'rules of the game' that filter access to administrative data for research purposes in the UK public sector? To address these questions, this work has treated the formation and prevalence of research data sharing institutions as processes consisting of social, professional and legal interactions between individuals who represent such organisational actors as data custodians, research groups and intermediary bodies. A qualitative empirical approach was adopted to study these processes. This approach relies on a constructivist-interpretative ontology which sees reality as socially

[224] N Waithira et al, 'Data management and sharing policy: the first step towards promoting data sharing' (2019) 17 BMC Medicine.

[225] *Infra*, chapter 6.

[226] E.g. Wellcome Trust, 'Policy and Advocacy team' <https://wellcome.ac.uk/about-us/teams/policy-and-advocacy-team>.

constructed and resists the idea of directly observable social-organisational structures.[227] Such structures, defined in this thesis as 'research data sharing institutions',[228] do exist and shape the decisions of individual civil servants, but are not static and rigid. They are not only comprehended, but also reconstructed and negotiated in everyday interactions.[229] Actors respond to 'changing conditions', competing over controlling the process of changing those conditions.[230]

Adopting this constructivist lens, the empirical chapters of this thesis present the way in which a particular regulatory space was constructed, reconstructed and recalibrated around the activity of research data sharing in the UK public sector over time. As the analysis will demonstrate, organisational actors have interacted in stabilising or renegotiating the institutions that govern research data sharing, e.g. an organisational policy framework, while constructing and consolidating their own organisational identities drawing on the meanings reflected in such institutions. For example, policy-makers consolidate a particular interpretation of the 'public interest' requirement through their precedent decisions, while at the same time drawing on the conventions followed within their organisational environment to position themselves towards the desirability of granting or refusing a request. In other words, data sharing institutions structure and influence individual action. At the same time, institutions are gradually reconstructed through the multiplicity of individual and organisational interactions in the research data sharing space.

By aiming to achieve an in-depth understanding of how data sharing decisions are shaped, this approach facilitates the iterative development of a well-integrated set of propositions,[231] which are to be tested in future research rather than generating a comprehensive, generalisable

[227] S Fuller, 'The Social Construction of Knowledge' in L McIntyre and A Rosenberg (eds), *The Routledge Companion to Philosophy of Social Science* (Routledge 2016) 358.

[228] *Supra* chapter 2, C on the definition of 'institutions' as a key dimension of this study's analytical framework, along with 'actors' and 'interactions'.

[229] P Ewick and S Silbey, *The Common Place of Law: Stories from Everyday Life* (University of Chicago Press 1998) 49.

[230] J Corbin and A Strauss, 'Grounded Theory Research: Procedures, Canons, and Evaluative Criteria' (1990) 13(1) Qualitative Sociology 3, 5; Dove (n 12) 96. See *infra,* chapter 4, A-C on the application of this in the UK research data sharing space.

[231] *Ibid.*

grand theory.[232] These propositions, thus, are not to be seen as positivist, nomological 'truths' about research data sharing but as a 'compelling understanding of the phenomenon' produced in relation to the theoretical assumptions adopted for this study, and 'against established standards of quality' in qualitative research.[233] The following section elaborates on how this approach was translated into the specific elements of my methodological research design.

B. Collecting Data about the Regulation of Research Data Sharing Decisions

I first discuss my choice of field setting and then my selection of data collection methods, as well as my sampling and interview preparation strategy.

i. Choice of Field Setting

To study the contemporary working and the historical development of the research data sharing space in the UK public sector,[234] empirical work mainly included data custodian bodies with experience in negotiating data access requests for social-scientific research purposes. Crucially, the enquiry was also extended to intermediary bodies and researcher groups which have been part of such negotiations. This was preferred to conducting a single case-study or multiple case-studies of particular data custodian bodies for a number of reasons.

First, while acknowledging that each data custodian may have their own particular response to research data sharing, it is the *activity* of such data sharing itself that gives rise to certain types of responses. As an activity which requires the weighing of risks and benefits, as well as the calculation of logistical and reputational costs, research data sharing induces reliance on organisational-level responses, drawing on cultural assumptions shared across specific organizations, e.g. large policy-making ministerial departments like the DWP and the HMRC, and governance structures, to justify the appropriateness of disclosure decisions. In that sense, deliberately selecting representatives from a wide number of data custodian bodies with some experience was preferable to an exhaustive consideration of specific bodies. As discussed in the previous section, the development of research data sharing has been incremental across

[232] S Hesse-Biber and P Leavy, *The Practice of Qualitative Research* (2nd ed, Sage 2011) 5.

[233] M Miles, M Huberman and J Saldana, *Qualitative Data Analysis: A Methods Sourcebook* (Sage 2013) 64.

[234] See *supra* chapter 2 p 33 on including a historical development perspective to the scope of the present study.

various public bodies, with small groups of civil servants within a given body being responsible for it. A selection of participants from a broader context was, thus, more important than including the insights of many officials within a single or a few bodies. The inclusion of representatives from key, 'pioneer' research data sharing bodies, which initiated this activity first and opened paths for other organisations, held greater potential to illuminate the formation of norms, ideals and practices governing this activity.[235]

Second, I extended the extension of the enquiry to intermediary bodies and requesting parties as necessary. Granting access to administrative data is a lengthy and technically complex process.[236] It involves various stages of communication between the decision-makers, the requesting parties and intermediary bodies aiming to facilitate data sharing. Exclusive focus on the custodian bodies would neglect the relational aspect of the data sharing process. It would also not allow the proper interrogation of a data custodian's construction of the reality of access decisions, by privileging their views compared to those of other actors.[237]

Furthermore, it was often the case that, due to the lack of resources and data-related expertise in the public sector,[238] individuals responsible for managing datasets on behalf of data custodians were seconded to different positions within the public sector. As a result, individuals working for such intermediary bodies as the Administrative Data Service (ADS) or the SAIL Databank would often be more experienced in negotiating data access for research purposes.

Finally, yet importantly, the bespoke nature of the research data sharing activity and the small number of civil servants and academics involved in it provided another reason against conducting case-studies. Had I conducted case-studies of particular custodian bodies and presented their specific attributes thoroughly, it would have been impossible to preserve the anonymity of my interviewees. Anonymity, however, was particularly crucial both for ethical reasons[239] and in the interest of securing access to interviewees, considering the sensitive nature

[235] The following sub-section provides more details on subject selection.

[236] E.g. Morris et al (n 116) mentions 'the timeline of 3 years and 4 months represents 6 face-to-face meetings and 108 email and telephone correspondences' before being granted access to data.

[237] Fuller (n 227).

[238] Veale, van Kleek and Binns (n 195).

[239] *Infra* B, iii on ethics.

of the discussed issues against the backdrop of criticism in relation to public-sector struggles to make better use of administrative data.[240]

ii. Selection of Data Collection Methods and Sampling

Interviews and documents, as well as informal conversations with research data sharing participants,[241] were used for collecting data about research data sharing decisions in the UK public-sector. Interviewing was employed to capture the thoughts and interpretations of research data sharing participants in respect of how the prevalent criteria and the appropriate requirements in the context of data access decisions have been shaped over time.[242] Semi-structured interviews were employed to both focus discussion and allow respondents to elaborate on the issues they considered most important.

The appendix provides a table with the interviewee codes, their affiliation and role within data sharing decision-making.[243] To provide some background to this study's interview sampling, forty-two (42) individuals were interviewed in total: twenty-two (22) individuals representing data custodians, seventeen (17) individuals from intermediary bodies and thirteen (13) individuals from major data-intensive research groups and organisations. As is implied in this breakdown of participants and explained in chapter 2,[244] some individuals participated while wearing two 'hats', i.e. representing both a data custodian and an intermediary body, or a research organisation and an intermediary body. In respect of job types, a relative majority (14/42) of the sample were policy-makers, with an almost equally significant part of the sample (12/42) being data analysts, and the rest (7/42) information governance officials.[245]

---

[240] *Supra* chapter 1, A.

[241] Mostly in the context of conferences which attract participation from both the research community and the civil sector, see *infra* (n 259). Note that these informal conversations were not strictly used as research data in this project, but informed the overall understanding of my research data.

[242] J Aberbach and B Rockman, 'Conducting and Coding Elite Interviews' (2002) 35(4) PS: Political Science & Politics 673.

[243] See appendices, C for the complete list of interviewees.

[244] *Supra* chapter 2 p 34-35.

[245] See *infra* chapter 8, C on the impact of this balance in participation on this study's findings. Future research should further probe the question of whether job role or organisational affiliation is more significant in shaping the 'rules of the game' in a regulatory space.

With data custodians, individuals working e.g. for such bodies as the Department for Work and Pensions (DWP), the Department for the Economy (NI) or the Knowledge and Analytical Services of the Welsh Government were interviewed. With intermediary bodies, interviewees worked e.g. for such bodies as the Administrative Data Service (ADS), the Economic and Social Research Council (ESRC) and the Wellcome Trust (WT). With requesting researchers, representatives of experienced groups in data-intensive social-scientific research based at such Universities as Queen's University Belfast, Essex and Swansea participated in the study.

Extracting data from documents was the other main data collection method used in this study to understand the efforts of data custodians and other stakeholders to establish norms and practices that govern research data sharing. Documents were used to collect information about all the dimensions of the research data sharing space: the way in which various organisational actors have expressed their aspirations to influence the creation of data sharing standards;[246] the way in which intermediary actors sought to lobby decision-makers in adopting researcher-friendly regulations,[247] as well as how researchers sought to communicate their concerns about the lack of transparency and clarity in data sharing procedures.[248]

Beyond the academic literature on public-sector data sharing,[249] several types of documents were used in this study. Legal documents, e.g. statutes and judicial decisions relevant to administrative data sharing for research, were included.[250] Websites were a key source of information in the fast-moving world of data sharing. For example, useful information could be drawn from the website of a data custodian, like the HBS in Northern Ireland,[251] outlining their data access requirements and the data sharing principles to which they subscribe. The websites of key intermediary bodies like the ADRN were also relevant when they would e.g.

[246] REFORM (n 117).

[247] J Rumbold and B Pierscionek, 'The Effect of the General Data Protection Regulation on Medical Research' (2017) 19(2) Journal of Medical Internet Research 47.

[248] Early Career Researchers Using Scottish Administrative Data, 'eCRUSADERS: About' <https://ecrusad.co.uk/>.

[249] E.g. Laurie and Stevens (n 7); Iveson and Deary (n 45); Jones et al (n 31).

[250] E.g. the Digital Economy Act 2017 or the UK Supreme Court judgment in *Christian Institute* (n 118).

[251] Health and Social Care Northern Ireland, 'Honest Broker Service' <http://www.hscbusiness.hscni.net/services/2454.htm>.

highlight the importance of particular legislative reforms in allowing a bigger scale of administrative data sharing for research purposes.[252] Minutes of meetings were another useful resource, particularly the ones of the ADRN Board meetings, convened by the UK Statistics Authority and attended by senior ADRN officers.[253] These documents often recorded deliberations and reflections on the appropriate requirements for researchers to get access to government data, e.g. by clearly demonstrating a connection between the research proposal and the 'public interest'.[254]

With both interviewee and document selection, a strategy of *purposive sampling* was adopted. Emphasis was put on selecting data sources that had the potential to yield insight about the various factors that shape data sharing decisions, according to my constantly developing explanatory account.[255] A 'theoretically meaningful' sample was sought by involving 'certain characteristics or criteria which helped develop and test' my emerging theoretical account.[256] In the case of documentary sources, I performed a targeted selection of documents by reference to, initially, broad directions in the data sharing literature and, then, my emerging findings about the regulation of research data sharing decisions.[257] The examples of relevant documents mentioned in the previous paragraph are indicative of this selection. In respect of interviews, I adopted a specific criterion of inclusion: substantive and long-term involvement in negotiating research data access, either on behalf of data custodians or on behalf of requesting or intermediary parties. Through negotiating access, individuals satisfying this criterion had contributed to the development of research data sharing institutions or adhered to such institutions.

In the absence of personal networks that would facilitate data access, as it often happens in elite interviewing,[258] I used a number of techniques to acquire access to the desired interview

[252] ESRC, 'Administrative Data Research Network' <https://esrc.ukri.org/research/our-research/administrative-data-research-network/>.

[253] E.g. UKSA (n 16).

[254] Office for National Statistics (n 150).

[255] D Silverman, *Interpreting Qualitative Data* (Sage 2015) 62.

[256] J Mason, *Qualitative Researching* (Sage 1996) 93-4.

[257] On the hypotheses, see *infra* C.

[258] Z van der Wal, 'Mandarins versus Machiavellians? On Differences between Work Motivations of Administrative and Political Elites' (2013) 73(5) Public Administration Review 749, 752.

targets. First, I attended and contributed twice to the annual Administrative Data Research conference, which brings together academics and civil servants engaged in administrative data sharing for research.[259] This conference proved a useful melting pot for various relevant stakeholders and allowed me to meet some of my future interviewees. Second, I had the benefit of relying on the help of a few 'insiders' that facilitated access to key interviewees.[260] These individuals were not government gatekeepers, but worked for intermediary bodies like the Administrative Data Service (ADS) or the SAIL Databank. Being well-versed in the practicalities of negotiating research data access, they were able to introduce me to their contacts in government and enable me to get valuable access.

Third, on a few occasions, I have used 'snowball sampling' to find new interviewees, either at periods when it was very difficult to access desired targets or when I felt that my interviewees could introduce me to other relevant, unknown to me, targets. In any case, even if I was introduced to someone, I did not conduct an interview without first establishing that the potential interviewee met my previously discussed involvement criterion.

iii. Interview Preparation and Ethics

Participants were recruited via e-mail.[261] A core interview guide was designed for the first eight (8) interviews and was then revised to reflect the progressive development of my explanatory framework. Slightly different interview guides were used for different interviewees to maximise the potential utility of the discussion. For example, senior or experienced staff were asked more questions about the original conception of data sharing practices and their development over time, whereas junior staff were asked more questions about their interaction with and interpretation of existing norms and practices.[262] Most interviews, i.e. twenty-five (25) out of fourty-two (42), were conducted face-to-face, with the remaining seventeen (17)

[259] 'The Digital Economy Act 2017 and its Conformity with Legal Restrictions on Administrative Data Sharing in the United Kingdom' (Administrative Data Research Conference 2017 – Edinburgh, June 2017); 'Custodian Legal Culture - Enabling Multi-Stakeholder Research Collaborations Between the Academia and Government in the UK' Abstract published in (2018) 3(2) *International Journal of Population Data Science* 8 (Administrative Data Research Conference 2018 – Belfast, June 2018).

[260] A Bryman, *Social Research Methods* (5th edn, 2016 OUP) 415.

[261] See appendices, A for an example of an e-mail recruitment message.

[262] An example is provided in appendices, D.

being conducted via Skype video call. Qualitative researchers have cautioned against the limitations of using Skype for interviewing due to difficulties in reading body language and nonverbal cues, as well as to the loss of intimacy compared to face-to-face interviews.[263] There is, however, a well-supported in the literature counter-argument that remote video conferencing software like Skype can be a valuable tool to complement or substitute face-to-face interviewing. Non-verbal cues and facial expressions can be identified via video call,[264] whereas not meeting in the physical working environment of the interviewee has been argued to 'increase presentation of self and authenticity compared with face-to-face interviews'.[265] In the present study, it is not expected that the mixed approach, comprising face-to-face and Skype interviews, has influenced the findings of the research. Interviewees consented to the use of Skype video call, which helped me identify non-verbal cues and facial expressions, and were very familiar with the use of the technology. In addition, using Skype allowed me to broaden the diversity of my sample by including individuals which I could not have interviewed face-to-face for logistical reasons.

Interview schedules, participant information sheets, recruitment letters and informed consent documents were reviewed and approved by the University of Oxford, Central University Research Ethics Committee.[266] Interviews were transcribed by me and securely stored in a password-protected hard drive. No individuals belonging to particularly vulnerable social groups were interviewed. Most of my interviewees were, in fact, well-educated and high-ranking civil servants or academics. Data were collected over a period of three years, from March 2017 to April 2020. Having discussed issues of collection, I now turn to data analysis.

[263] S Seitz, 'Pixilated partnerships, overcoming obstacles in qualitative interviews via Skype: a research note' (2016) 16(2) Qualitative Research 229-235; N King, C Horrocks and J Brooks, *Interviews in Qualitative Research* (2nd edn, Sage 2019) 121.

[264] V Lo Iacono, P Symonds and D Brown, 'Skype as a Tool for Qualitative Research Interviews' (2016) 21(2) Sociological Research Online 103.

[265] R Janghorban, R Latifnejad Roudsari and A Taghipour, 'Skype interviewing: The new generation of online synchronous interview in qualitative research' (2014) 9(1) International Journal of Qualitative Studies on Health and Well-being.

[266] Ref No: R50823.

C. An Issue-Focused, Thematic Analysis of the Regulation of Research Data Sharing Decisions

This work initially involved an exploratory stage, identifying candidate factors which potentially explain the formation of research data sharing institutions by organisational actors in the UK public sector. After identifying a set of factors, the research engaged in explanatory work, seeking to refine them, as well as develop a well-integrated set of propositions about the relationships between them and their relative salience in regulating data sharing decisions. During data analysis, it became evident that three particular dimensions of this study's analytical framework – 'actors', 'institutions' and 'interactions' – could provide a meaningful structure within which meaningful categories from the data would be situated. This is reflected in the following chapters of the thesis.[267]

In fleshing out these dimensions of the research data sharing space, data analysis was *issue-focused* rather than case-focused.[268] In other words, analysis focused on the activity of research data sharing as an overarching 'issue' triggering similar response across the board, rather than on the particular responses of specific 'cases' of organisations in this space. The aim of this strategy was to strike a balance between the general and the contextual by seeking to portray 'what has been learned from all respondents about people in their situation'.[269] Data analysis started after the completion of the first round of conducting semi-structured interviews in England.[270] At this stage, document data involved the selection of a first set of ca. twenty (20) documents, *inter alia* minutes of the ADRN board meetings,[271] public-sector data sharing review reports,[272] websites of key organisations[273] and other documents produced by data

[267] *Infra* chapter 4 presents findings about the actors who have competed for parts of the research data sharing space over time; chapters 5 and 6 present findings about the institutions or 'rules of the game' that filter access to administrative data and chapter 7 elaborates on the way in which interactions between participants of the space have been central in the above developments.

[268] R Weiss, *Learning from Strangers: The Art and Methods of Qualitative Interview Studies* (The Free Press 1994) 152.

[269] *Ibid* 153.

[270] By that time, eight (8) individuals from two data custodians (the Department for Work and Pensions and the Office for National Statistics) and one intermediary body (the Administrative Data Service) had been interviewed.

[271] UKSA (n 16).

[272] Research Project Management (n 31); Thomas and Walport (n 2); ADT (n 25).

[273] E.g. ESRC (n 252).

custodians that were available in the public domain.[274] With both interview transcripts and documents, the NVivo 12 software was used to manage and analyse the data by 'marking passages of text in a project's documents with nodes'.[275]

In analysing the data, a 'thematic analysis' approach was followed, aiming to 'identify, analyse and report patterns within data'.[276] This was an *inductive* and *grounded* approach to analysis, identifying patterns from the data in a bottom-up way, even if I did not follow all the procedures of a grounded theory approach. For example, I coded sentences or small paragraphs in lieu of performing exhaustive line-by-line coding. I did also not deliberately disregard the existing literature on the potential explanatory factors of the phenomenon.[277] Such literature provided me with broad directions in the early phases of coding. Analysis involved a succession of stages. Data were read and familiarised with, taking note of items of potential interest to answer the research question by creating the relevant 'nodes' in NVivo. In the first instance, this involved creating first-order, semantic 'codes' which could be associated with the formation of institutions governing access decisions. Such codes could have to do with the law, e.g. with legal provisions or 'gateways', or with the type of research as more 'applied' or 'theoretical'. For example, this might involve an interviewee comment that 'the lack of a permissive legal gateway is prohibitive for us when dealing with a request' or that the data custodian body they represent is not interested in 'blue skies research'.[278] Similarly, I would code an excerpt from a document produced by a data custodian which would mention that they take seriously the fact that their legal powers to share data for research purposes are limited.[279] At this time, I would keep a memo of my impressions about the meaning and the significance of these first-order codes within my interviewees' accounts.

[274] J Littlechild, 'Developments in access to DWP data' (2015) <https://esrc.ukri.org/files/research/administrative-data-taskforce-adt/developments-in-access-to-dwp-data-joanna-littlechild/>.

[275] Bryman (n 260) 606.

[276] V Braun and V Clarke, 'Using Thematic Analysis in Psychology' (2006) 3 Qualitative Research in Psychology 77; Dove (n 12) 72, 198.

[277] Silverman (n 255) 119.

[278] Interview with participant nr. 19 (2018, Swindon) meaning research that is too theoretical.

[279] Littlechild (n 274).

These codes were then grouped together under second-order, hierarchical nodes, which indicated the emergence of particular 'themes' in the data. One theme, for example, was that 'research not linked to policy-making is not worthwhile'. This was a researcher-derived theme, within which I coded statements like the previous one on the low value of research that is not relevant to policy-making priorities, as well as e.g. organisational documents discouraging requesting researchers from submitting requests based on high-level or theoretical projects.[280] Another set of themes, with a focus on the actors-participants of the space, had to do with the type of requesting research organisations at particular points in time. One theme started to emerge from the association of multiple first-order nodes related to 'commissioned research by the government for technical issues', whereas another theme referred to 'ad-hoc research by trusted academics'.

After the creation of a number of 'themes', I started identifying particular 'patterns' that explain the relationship between the themes by reference to recurrent features in the data.[281] The theme of 'research not linked to policy-making is not worthwhile' was relevant to the opposite in meaning theme of 'research directly relevant to policy-making is very important'. The relevance of the themes to each other would be made clear when interpreting the narratives of my interviewees as a whole. For instance, an interviewee working for a data custodian would mention that even if they were not extremely happy with a particular group of requesting researchers from previous collaborations, they would still amend favourably their data sharing procedure for them 'if we can agree on the research questions'.[282] Relying on my interviewees' accounts about the importance of these themes within the data sharing process, the pattern of a 'request's relevance to policy-making justifies data access' started to emerge. Similarly, interpreting the previously mentioned themes of 'commissioned' and 'ad-hoc' research in the light of other relevant themes, the pattern of 'informality' as a distinct historical phase in the development of the research data sharing space emerged.[283]

---

[280] Office for National Statistics (n 150).

[281] J Ritchie and L Spencer, 'Qualitative data analysis for applied policy research' in A Huberman and M Miles (eds), *The qualitative researcher's companion* (Sage 2002) 310.

[282] Interview with participant nr. 16 (2018, phone).

[283] See *infra*, chapter 4, A on 'informality' as one of the phases in the historical development of the UK research data sharing space where norms and practices within informal relationships between data custodians and researchers by and large governed access to data.

In developing themes into patterns, I used the 'constant comparative method'.[284] During this exercise, I aimed to include settings and subjects which could modify or broaden the initial findings, e.g. by including deviant cases that could challenge my initial hypotheses. Existing themes and patterns were refined by merging similar entries, discarding the ones which became less relevant and creating new ones. To give an example, it was initially assumed that restrictive interpretations of the law are associated with the prevalence of stricter and more restrictive data sharing frameworks on-the-ground. Yet, it seemed that such custodian bodies as the Honest Broker Service (HBS) in Northern Ireland were both adopting restrictive legal interpretations and working within a friendly to disclosure decision-making framework. This was confirmed through close inspection of the documents they produced, as well as through an in-depth interview with three of the individuals managing the HBS datasets.[285] Furthermore, the relationships between different patterns was clarified through implicitly prompting my interviewees to assess their relative strength, both within qualitative data collection and within a test survey designed to enhance the validity of this study's findings.[286] For instance, individuals working for the Knowledge and Analytical Services in the Welsh Government suggested that the legal framework is not facilitative of research data sharing, yet insisted that it is not 'more law' that is central to streamlining data sharing procedures but 'better relationships' and trust.[287]

Ultimately, a framework of patterns was developed within the three analytical dimensions. With *actors*, the developed patterns suggested the existence of interconnected historical phases of development of the space, with particular actors invoking certain narratives that marked crucial transitions. With *institutions*, patterns reflected the existence of such driving factors as the legalities and technicalities of data sharing or the relevance of research to policy-making as relevant to the establishment of data sharing frameworks. With *interactions*, emerging patterns suggested that low-level communications are more salient in shaping the boundaries of the space than high-level policy decisions. This set of propositions support the central argument advanced in this thesis, i.e. that commonplace interactions, informed by underlying

[284] K Charmaz and A Bryant, 'Grounded theory and credibility' in D Silverman (ed), *Qualitative Research* (3rd edn, Sage 2011) 292.

[285] Interview with participants nrs. 11, 12 and 13 (2018, Belfast).

[286] *Infra* D, i.

[287] Interview with participants nr. 25 and 26 (2019, Cardiff).

policy priorities, between data-savvy analysts and researchers, rather than momentous high-level policy decisions have been constantly reshaping the regulation of research data sharing decisions in the UK public sector. Chapters 4-7 elaborate on this argument and its supporting propositions. Empirical work reached its conclusion when fresh data no longer produced new insights.

D. 'Validity', 'Reliability' and Limitations

i. 'Validity'

Within qualitative research designs, validity is associated with 'credibility' or 'correspondence between observations and theoretical ideas'.[288] To achieve such credibility, qualitative researchers often provide an account of their findings to the people about whom the research has been conducted for assessment purposes.[289] A main challenge in performing this exercise is that participants cannot validate a researcher's *analysis*. By and large, they can only express a view on whether they agree with the main conclusion or whether their views have been accurately registered by the researcher.[290]

For the purposes of a relevant assessment, I designed and implemented a pilot survey to a small convenience sample (n=39) of research participants in a conference about administrative data sharing for research. Respondents to the survey included both some of my interviewees and other representatives from data custodians, such as the Department for Education (DfE) in England or National Records Scotland (NRS), as well as intermediary bodies (e.g. the Administrative Data Research Centre-Scotland) and research groups (e.g. the Population Data Science group at Swansea University). The pilot survey was designed by reference to my qualitative findings about the factors that structure and justify research data sharing

[288] Mason (n 256) 24.

[289] V Braun and V Clarke, *Successful Qualitative Research* (Sage 2013) 282.

[290] See B Skeggs, 'Situating the Production of Feminist Ethnography' in M Maynard and J Purvis (eds), *Researching Women's Lives from a Feminist Perspective* (Taylor & Francis 1994) 86: 'the most common response of the respondents was ''can't understand a bloody word it says'' when they were fed back some of the writings'.

decisions.[291] The different factors were translated into Likert[292] items through accessible questions about organisational rationales and custodian criteria behind research data sharing.[293] This allowed participants to directly engage with my theoretical categories, as well as express their views on their validity and relative strength.

To further facilitate this engagement, I included an open-text box in the survey, urging interviewees to write 'in case the above categories are not applicable to your experience or you have additional comments to describe it'.[294] After the survey, I liaised with some of the participants, including previous interviewees of mine, prompting them to comment on the validity of the identified in the survey factors. I tried to build rapport with all participants and hear carefully their views. They were broadly positive about the correspondence of the key theoretical categories with their experience, mainly prompting me to provide them with more details about my explanation or dig deeper into certain areas. For example, I was prompted to assess the impact of pre-existing relationships between requesting academics and civil servants responsible for granting access on the development of data sharing frameworks that are likely to allow wider data disclosure. I pursued this during the last phase of my qualitative data collection, in the interest of enriching my understanding of the interactions between the different actors in the research data sharing space.

ii. 'Reliability'

Reliability is defined within qualitative research designs as 'dependability', mostly associated with following a transparent, reflective and clearly defined research process.[295] In other words, the researcher should give assurance to external audiences that a high-quality, systematic piece of research was conducted in good faith.[296] Systematic literature reviews of qualitative studies in public administration research indicate a 'lack of transparency of the path from data

[291] *Supra* C on the production of the hypotheses.

[292] A Likert scale is a multiple-indicator measure of a set of attitudes relating to a particular area, aiming to measure agreements and disagreements on factors relevant to the area in question, see D de Vaus, *Surveys in Social Research* (5th edn, Routledge 2001) 182. For examples of such questions, see appendices, E.

[293] Appendices, E.

[294] *Ibid*.

[295] Y Lincoln and E Guba, *Naturalistic Inquiry* (Sage 1985) 316.

[296] J Moisander and A Valtonen, *Qualitative Marketing Research: A Cultural Approach* (Sage 2006) 27.

collection to interpretation and theorising' as the major 'trustworthiness' concern in the field.[297] The present chapter has aspired to present transparently the way in which qualitative data were collected, analysed and interpreted in the present study.

Apart from that, particular steps were adopted to mitigate risks of *bias* in the work. As mentioned, I kept detailed records and memos during data collection and analysis, transcribing all the interviews fully to ensure an accurate representation of what the participants said.[298] The same sampling criterion was consistently applied in data collection,[299] consulting key stakeholders and my 'insiders' about its application throughout the research. All available and pertinent data sources were used. Beyond feedback received through supervision, drafts were subjected to peer review at various stages of the work, either in informal workshops (e.g. at the CSLS and at HeLEX) or in conferences (e.g. the ADR Conference or the SLSA).

iii. Limitations

Methodological limitations in conducting this research have mainly stemmed from difficulties related to this not being a case-focused analysis of particular organisational actors. This also had to do with the particularities of the research population of data custodians in the UK. As previously discussed,[300] data custodians are far from a clearly demarcated and well-reported about population. Even if representativeness was not the purpose of my purposive sampling and issue-focused data analysis,[301] a broad selection was desirable to mitigate bias risks and enhance the quality of the findings.[302] Hence, I sought to include interviewees from different types of public bodies across the four UK nations (e.g. ministerial departments, health boards

[297] S Ospina, M Esteve and S Lee, 'Assessing Qualitative Studies in Public Administration Research' (2017) 78(4) Public Administration Review 593, 599; R Brower, M Abolafia and J Carr, 'On Improving Qualitative Methods in Public Administration Research' (2000) 32(4) Administration & Society 363.

[298] K Charmaz, *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis* (Sage 2006) 3; a similar technique was adopted by Dove (n 12) 201.

[299] *Supra* B, ii.

[300] *Supra* A.

[301] *Supra* B, ii and C.

[302] Bryman (n 260) 416.

or agencies), as well as individuals working in different layers in decision-making (e.g. policy-makers, information governance officers or data analysts).[303]

Maximising range and width, however, was challenging due to the general nervousness around digital experimentation in the public-sector,[304] as well as the particular criticism around public-sector struggles to use administrative data to improve policy-making.[305] A small number of my requests (ca. 5-7) to relevant interview targets were rejected or never answered, whereas others replied that they do not have relevant experience. As a result, while observing my criterion of experience and involvement in data access negotiations, I have not managed to interview data custodians from some types of bodies e.g. local authorities. This may have created limitations for my theoretical understanding of the formation of data sharing institutions in the public-sector by e.g. failing to reflect a particular dimension of data sharing culture that is associated with local-level policy-making. Similar statements can be made about other types of bodies not included in this study e.g. financial or environmental public authorities. For example, bodies holding financial data may perceive themselves as prone to heightened confidentiality risks, whereas environmental agencies might be keener to withhold information because they see them as their leverage in regulating their market.[306]

Another way in which sampling difficulties may have influenced the findings of this study relate to the underrepresentation of information governance officials within my sample. It is, indeed, possible that this may have influenced my interpretation of the technicalities and legalities of data sharing as less determinative of the actual creation of research data sharing frameworks compared to policy considerations. To mitigate this limitation, I made sure that my interview guides include all the categories of my emerging theoretical account and not only the ones that are most relevant to my specific interviewee, e.g. not asking a policy-maker only about policy considerations. I also used my alternative data sources, documents and off-the-record conversations with data sharing participants, to constantly interrogate the interpretation of my interview data.

---

[303] Appendices, C.

[304] Booth (n 64).

[305] *Supra* chapter 1, A.

[306] B Lange, 'Compliance Construction in the Context of Environmental Regulation' (1999) 8(4) Social & Legal Studies 549.

Furthermore, due to the necessarily selective choice of interviewees and the decision not to conduct particular case studies,[307] the discussion of the 'research data sharing decision-making process' as the unit of analysis operates at a level of significant abstraction. While nuances in different organisational environments may exist, this study aimed to capture the way in which the activity of research data sharing has triggered the same type of responses across the UK public sector. In that regard, representatives from key, pioneer data sharing bodies were prioritised as more likely to have influenced the development of data sharing norms and practices across the UK public sector. Nonetheless, my findings may not reflect the way in which data sharing institutions are shaped in organisational environments different from the ones presently included in the study.

In spite of the limitations, the present analysis has attempted to navigate a sensitive and complex empirical field, generating a well-integrated set of propositions about the formation of research data sharing institutions in the UK public sector. It is hoped that future research can test, refine or revise these argumentative propositions,[308] particularly when more comprehensive data on access decisions will become available by the UK government. The following chapters present these arguments and discuss the empirical data on which they are based.

[307] *Supra* B, i.

[308] *Supra* chapter 1, B.

4. BETWEEN 'EFFICIENCY' AND 'ACCOUNTABILITY': THE ACTORS WHO SHAPED THE HISTORICAL DEVELOPMENT OF ADMINISTRATIVE DATA SHARING FOR RESEARCH IN THE UK PUBLIC SECTOR

> *'All those involved, from data holders to the general public, (now) have a clearer understanding of the needs of the research community and researchers can plan their research and be confident that access to administrative data will be granted.'* [Elias, *The UK ADRN: Its Genesis, Progress and Future*]

> *'The landscape of administrative data (...) is complex and changeable, and is often difficult for researchers to navigate. Problems mainly relate to the length and complexity of the process – making for a challenging research environment.'* [Iveson and Deary, *Navigating the Landscape of Non-Health Administrative Data in Scotland: a Researcher's Narrative*]

> *'I think there has to be the aspiration and the intention that this is public data. (...) We'd like to see a time when the academic researcher will say I'm asking these questions, because we should get the answers, and that is not challenged by the Department.'* [*Interviewee 10*]

The regulatory space lens is apt for understanding power struggles between the actors populating the UK research data sharing space. Through this lens, this chapter takes stock of the historical development of UK government research data sharing in the last four decades (1980-2020). It does so to analyse how processes of 'conflict' and 'bargain' between data custodians, researchers and intermediary or high-level standard-setting bodies recalibrated this particular space, shifting time and again the goalposts when it comes to appropriate uses of government administrative data for social-scientific research purposes. More specifically, this chapter sketches the way in which particular organisations have defined the scope of central issues during the development of this space over the last four decades.

In doing so, this chapter draws on documentary data, such as legal and policy documents enacting research data sharing frameworks, and data from fourty-two (42) semi-structured interviews with individuals working for organisations with experience in research data sharing.

Actors in the UK research data sharing space, it is suggested, shaped the historical development of this activity over three, broadly constructed, phases: '*informality*', '*formalisation*' and '*fragmentation*'. In the first phase, which spans from the 1980s until 2004, there were no formal procedures for researchers to be granted access to government data. The success of relevant efforts largely relied on ad-hoc bargain, through the personal relationships between such actors as requesting parties and data custodians or government sponsorship of large research projects.

In the second phase, from 2004 to 2018, actors on all sides of the playing field invoked strong narratives about widening access to administrative data for research purposes. These narratives, often voiced in a bold and demanding manner within a conflict between different organisations, triggered the creation of formal procedures and frameworks for allowing research data sharing to accredited researchers. Finally, in the third phase, which is currently underway since 2018, it becomes apparent that researchers' struggles to demand a more formalised research access to government data have not materialised in consistent and predictable frameworks across the UK. While some formal frameworks exist, it is often the case that many data custodians and intermediary bodies do not publicise any or much detail about their data sharing procedures and supported types of research, leaving researchers in the dark about their precise requirements or their interpretation of crucial aspects of existing frameworks. A picture of fragmentation emerges, with researchers often facing insurmountable barriers to their efforts to get access in a timely and satisfactory manner, contrary to assertions that researchers can 'plan their research and be confident that access to administrative data will be granted'.[309] Some researchers adopt a more adversarial approach to navigating this landscape, whereas others prefer to bargain with data custodians, much like their counterparts in the early phases of UK research data sharing.

The three phases identified in this chapter are naturally fuzzy, with no clear-cut divisions from one phase to another, so that it is not possible to categorically pin down particular moments in time when one phase fully 'ends' and the succeeding phase 'begins'. In that sense, for example, it is not claimed that the phase of 'fragmentation' starting in 2018 and lasting until today is not unique, since there has always been fragmentation in the space as the empirical data will demonstrate. Nonetheless, the presentation of the identified phases and the ways in

[309] P Elias, 'The UK Administrative Data Research Network: Its genesis, progress, and future' (2018) 675(1) The Annals of the American Academy of Political and Social Science 184.

which one of them gives rise to another is analytically meaningful not only because it illuminates the historical path dependencies that triggered the historical development of the space, but also because the construction of the phases and the transitions between them reflect the views of the participants of the space about the regulation of data access. To use the example of fragmentation again, the label of 'fragmentation' is indicative of the tensions between literature assertions that routes to data access for researchers are now transparent and well-structured[310] and interview data demonstrating that the landscape is actually quite patchy and access to data is unequal and unpredictable.

Two central narratives stand out and determine who is 'in' and who is 'out' of the research data sharing space: '*efficiency*', i.e. the idea of facilitating research data sharing with a view to create an evidence-basis for better policy-making, and '*accountability'*, i.e. the idea of facilitating research data sharing with a view to stimulate critical engagement with the government's policies in the name of transparency and openness. While most actors in the space have paid lip service to the compatibility of the two narratives, the analysis demonstrates that 'accountability' has been gradually eclipsed by 'efficiency'. The role of particular data custodians, mostly large ministerial departments like the Department for Work and Pensions and the Department for Education, as well as intermediary bodies like the Administrative Data Research Network, has been central in that process. This is not irrelevant to broader, systemic-level changes in the UK public sector over the last four decades. New public management rationales and neoliberal approaches to managing the civil service promote certain solutions prevalent in the UK research data sharing space, e.g. business-like resource management, contracting-out services like research, cutting down investments to 'do more with less',[311] over traditional accountability, rule-of-law concerns. I first analyse how actors invoked the two narratives, 'efficiency' and 'accountability', to structure the three historical phases of UK research data sharing and triggered transitions from one to another, before elaborating on the main questions arising for the formation of research data sharing institutions in the UK public sector.

[310] *Ibid.*

[311] E Ferlie et al, *The New Public Management in Action* (OUP 1996) 6; also see Dove (n 11) 70 on the 'neoliberal discourse stressing market rationality and economic optimisation' underpinning health research regulatory reform in the UK.

A. Informality: Ad-Hoc Bargaining to Access Data for Research

The origins of social-science research data sharing in the UK were fraught with technical difficulties and data custodian reluctance to take risks. Departments, by and large, commissioned research for their own purposes and, incrementally, started to engage in ad-hoc arrangements with trusted external academic researchers for independent research projects.[312] As it will be shown, through adopting both bargaining and conflicting strategies, researchers instigated a transition from the informal governance of data access towards more formalised procedures and data sharing frameworks.

i. Controlled access for researchers and data custodian reluctance

In the 1960s, administrative records were collected for the first time at a massive scale in the UK and information about them was disseminated via official statistical gazettes. At that time, allowing access to these records for research purposes was not deemed as legitimate.[313] In the late 1970s, however, the UK government started to appreciate the 'research value inherent within their large and growing databases'.[314] Departments started producing research in the interest of informing policy delivery, by and large commissioning professional researchers to work on research questions of interest to the department or having in-house statisticians to analyse the datasets.[315] The 1980s brought a revolution in computerisation and, through that, more systematicity and standardisation in collecting administrative records and using them for research purposes. In the late 1980s and 1990s, departments started to use administrative data for research purposes on a routine basis.[316]

[312] This is to distinguish commissioned research from independent research, in the sense that commissioned research projects have pre-determined research questions and (often) methods, with the researcher having limited scope as to following a different course of action during the project.

[313] Elias (n 303) 185.

[314] *Ibid* 186.

[315] P Jackson, 'From ''Intruders' to 'Partners'': The Evolution of the Relationship Between the Research Community and Sources of Official Administrative Data' in N Crato and P Paruolo, *Data-Driven Policy Impact Evaluation* (Springer 2019) 17.

[316] Elias (n 303) 186.

While this time also saw the establishment of the first research data centre for health data in the UK by the Office for National Statistics and the Economic and Social Research Council to stimulate the interest of external researchers in analysing government data,[317] similar developments in social-science research data took longer to materialise. Access to government data for social research purposes was generally envisaged as *controlled* (or restricted), as opposed to *registered*.[318] In the case of *controlled* access researchers who satisfy certain technical, legal and ethical requirements apply for access on a *project-by-project* basis. Their applications, often accompanied by comprehensive research proposals, are reviewed by sector-specific committees or by the data custodians themselves.[319] *Registered* access, on the other hand, envisages the provision of access to users by reference to their *credentials* (e.g. researchers affiliated with recognised institutions or complying with specific accreditation requirements), rather than requiring a case-by-case decision on the merits of a specifically described research project.[320]

Data custodians have been reluctant to widen access to administrative records for social research purposes. With administrative data effectively being treated as property of public bodies,[321] *controlled* access has been and currently is the dominant model in the UK public sector. This is the case for a number of reasons. Firstly, risks related to the disclosure of identifiable data is deemed as one of the main reasons behind restricted access. Perceptions of such risks were exacerbated in the advent of data breach incidents involving the loss of millions of personal complete data records, putting pressure on data custodians to minimise the risk of

[317] F Ritchie, 'Secure access to confidential microdata: four years of the Virtual Microdata Laboratory' (2008) 2 Economic & Labour Market Review 29.

[318] S O Dyke et al, 'Registered access: authorizing data access' (2018) (26) European Journal of Human Genetics 1721.

[319]E M Ramos et al, 'A mechanism for controlled access to GWAS data: experience of the GAIN Data Access Committee' (2013) 92 American Journal of Human Genetics 479–88.

[320] S O Dyke et al, 'Registered access: a ''Triple-A'' approach' (2016) 24 European Journal of Human Genetics 1676–80.

[321] Although, legally speaking, they are not their 'property' according to a consistent line of UK Court of Appeal case law that shows that right-holders in respect of data do not have property in the data: *Fairstar Heavy Transport NV v Adkins* (2013) EWCA Civ 886 (Mummery LJ) [46-50]; *Your Response Ltd v Datateam Business Media Ltd* (2014) EWCA Civ 281 (Moore-Bick LJ) [29-34]; *Environment Agency v Churngold Recycling Ltd* (2014) EWCA Civ 909; as well as the Chancery Division in *Capita Plc v Darch* (2017) EWHC 1248 (Ch). Chapter 5, *infra* elaborates on the legal framework that applies to research data sharing.

such leaks in the future.[322] An interviewee who has worked for the Department for Work and Pensions recalls the nervousness caused within the department after a massive data breach concerning child benefit data:

> *'It was always quite difficult to get hold of the benefit data. After the loss, everything came to an absolute standstill for a year and nobody could get access to it. Even by the time I left, you had to be next to God before you could get access to data.'*[323]

As another interviewee, who has worked for the Department for Education, explains, such disclosure concerns may be traditionally stronger at particular departments like the Department for Work and Pensions (DWP) or HM Revenue and Customs. In their words, disclosing people's earnings and benefits is a 'tough thing for a government department to do'.[324] Data custodians are likely to fight tooth and nail to prevent this from happening. Two interviewees currently working for the DWP corroborate this assertion. They attribute their department's cautious attitude towards disclosing data to researchers to their need to 'gain assurance that a lot of the things we've got confidence in internally have been applied externally',[325] making sure that requesters do not 'ask for a huge amount of detail'[326] that would exacerbate identifiability risks.

Secondly, government departments have allegedly been perceiving external researchers as less responsible data processors compared to their own statisticians,[327] preferring to monitor each and every disclosure for research purposes. An interviewee who has worked for the Administrative Data Service, an intermediary body seeking to facilitate wider access to administrative data for academic researchers, recalls that the DWP's 'established processes for

[322] Administrative Data Taskforce (n 23) 23.

[323] Interview with participant nr. 30 (2020, virtual).

[324] Interview with participant nr. 33 (2020, virtual).

[325] Interview with participant nr. 8 (2017, Sheffield).

[326] Interview with participant nr. 7 (2017, Sheffield).

[327] Jackson (n 308): 'Social and economic research in academia, which had most to gain from using the personal digital information held by public authorities, did not progress in a coherent and deliberate way with reforms that mirrored the changes the official data community were undergoing'.

receiving requests for data' had to be re-designed before applying to academic requesters.[328] According to them, this might necessitate protracted processes of bargain between the requesting party and the data custodian, with the latter insisting on tangible proof that the former are to be trusted as reliable data users. Researchers need to demonstrate not only their skills and competences as professionals, but also their patience and charm as individuals to move things forward during these long-drawn bargaining.

Thirdly, investing into wider and systematic research access to government data would require significant investment from a logistical perspective.[329] The Office for National Statistics (ONS), for example, have been finding the cost of providing remote access to their Virtual Microdata Laboratory (VML) for the academic community to be 'prohibitively expensive'.[330] Crucially, departments have struggled to see the economic benefits in widening access to their data if they are not the 'beneficiaries of the secondary analysis' since the benefits to the public interest of the society as a whole generated by academic social research are 'diffuse (…) (and) difficult to quantify'.[331] An interviewee doing administrative data research in Scotland suggests that, in many cases, reluctance to establish a consistent data sharing framework is caused by high turnover of staff or the lack of resources to recruit a 'decent-size' team.[332] Other interviewees, working for the Economic and Social Research Council and the ONS, confirm that reluctance to widen access to administrative data is often related to departments being under-resourced.[333]

Fourthly, and finally, reluctance is often a by-product of organisational inertia, in cases where a type of technology such as remote data access or data linkage has not been widespread enough for decision-makers to be familiar and comfortable with it. An interviewee who has worked for the ONS at the time of the establishment of its VML data infrastructure vividly recounts how they convinced the decision-making board to endorse their proposals to establish

[328] Interview with participant nr. 6 (2018, phone).

[329] T Desai, F Ritchie and R Welpton, 'Five Safes: designing data access for research' (UWE Economics Working Paper Series) (2016) 3.

[330] F Ritchie, 'UK release practices for official microdata' (2010) 26(3) Statistical Journal of the IAOS 103.

[331] F Ritchie and R Welpton, 'Addressing the human factor in data access: incentive compatibility, legitimacy and cost-effectiveness in public data resources' (UWE Economics Working Paper Series) (2014) 4.

[332] Interview with participant nr. 29 (2020, phone).

[333] Interviews with participant nr. 19 (2018, Swindon) and 28 (2020, London).

a novel data sharing procedure by mentioning another existing application of a similar framework:

> *'When I first went to the ONS board (…) presenting the idea for the VML (…) we were, just, not going anywhere. The board was sitting around and I could see they are unhappy about this stuff and they are not going to sign up for this. And then, someone asks and says, you know, is this the first time this technology has been employed? And I said, no, Denmark's already got this system. And they seemed very happy with it! And that was it. Everyone said, alright, that's ok then. We don't care as long as we are not the first, we don't want to take the risk.* '[334]

Delicate bargaining like the one described in this case was of the utmost importance to allay custodian worries. With these considerations in mind, it is hardly surprising that data custodians fought hard to confine the first administrative data research projects in the UK public sector within their organisational boundaries.

ii. The first administrative data research projects: doing research in-house or commissioning academic researchers

As data custodians started collecting and curating new and large datasets with 'direct relevance to policy issues' in the advent of digitalising government services,[335] both central and local government across the country explored the potential of administrative data research to inform policy-making. In the late 1980s and the 1990s, government departments started investing into administrative data research in two ways: either by setting up teams with computer science and statistical expertise to conduct research in-house, or by 'contracting-out' research to academic researchers with relevant experience.[336]

[334] Interview with participant nr. 42 (2020, virtual).

[335] G Smith et al, 'The Value of Linked Administrative Records for Longitudinal Analysis' (ESRC National Longitudinal Strategy Committee) (2004) 5.

[336] Elias (n 303) 186. Non-academic researchers, such as private sector consultants, were also contracted out during this period, but this is outside of the scope of this thesis as explained *supra* chapter 2 p 30.

Certain data custodians like the Department for the Economy (DfE), the Department for Work and Pensions and the Office for National Statistics were the pioneers in promoting research data sharing initiatives. Two interviewees who have worked for the DfE offer some background on the emergence of research data sharing within their department. According to their recollections, most of the social research was initially happening within the department. The 1990s, however, saw a 'big push to contract out as much as possible' in the public sector, with social research being close to the top of the list of the things that would be contracted out.[337] Beyond political drives to reduce permanently employed public sector staff in the context of austerity and new public management,[338] there were other reasons for contracting out administrative data research. In some cases, such research may involve complex and sophisticated statistical methods, which the department 'wouldn't have the capacity to do'.[339] In other cases, the department would find it more palatable to present a particular research finding as carried out by someone 'neutral and independent' rather than by its own statisticians.[340]

When contracting-out a research project, a government department would define the question that would be answered, the variables the researcher would use and the 'hypothesis that they would support'.[341] As an interviewee, who has undertaken commissioned research for the DfE recalls, apart from agreeing on the research objectives, there would be little in terms of data sharing processes at the time:

> *'If we go back right at the start, there were not really any processes. You just said I want to use the data for this purpose and it would be sent to you via a CD. Thinking about that now, that's mad, really! But that's how it was. Over time, processes have become more formalised and there are more hurdles to jump over.'*[342]

[337] Interview with participant nr. 30 (2020, virtual).

[338] See K McLaughlin et al (eds), *New Public Management: Current Trends and Future Prospects* (Routledge 2005).

[339] Interview with participant nr. 33 (2020, virtual).

[340] Interview with participant nr. 30 (2020, virtual).

[341] Interview with participant nr. 10 (2018, Belfast).

[342] Interview with participant nr. 40 (2020, virtual).

As this description suggests, at this stage and at least in some examples of data sharing, it might even have been the case that researchers were able to negotiate the scope of their aspired data use without ruffling too many feathers within data custodian bodies. While departments like the DfE may have occasionally commissioned external researchers to perform analysis on their datasets, the most active department in maximising the use of its administrative data for research purposes in England is certainly the Department for Work and Pensions (DWP). As an interviewee who has worked for the Administrative Data Service explains, the DWP has collaborated with academic researchers 'very closely within specific policy issues'.[343] Beyond contracting academics for the purposes of conducting government research, however, the Department has also invested significantly in building multi-disciplinary teams of data analysts who collaborate with policy officials in the interest of supporting the latter with an evidence-basis for and evaluation of the department's policies.[344]

Major administrative data research projects have been conducted by the DWP. The Work and Pensions Longitudinal Study is an 'unmatched' longitudinal data infrastructure, with complete coverage of unemployment benefit claimants,[345] that has been used for a variety of policy purposes e.g. counter-fraud or evaluating the success of Jobcentre plus.[346] Through using insights generated from this study, the DWP has been seeking to maximise its 'operational effectiveness' and support its agenda in key policy areas such as child poverty, welfare-to-work and retirement planning by targeting resources to the appropriate people.[347] The 'new deal for lone parent evaluation' is another example of a major research project, within which the DWP contracted experts from the National Centre for Social Research to analyse

[343] Interview with participant nr. 5 (2017, Essex).

[344] Department for Work and Pensions, 'Research at DWP' <https://www.gov.uk/government/organisations/department-for-work-pensions/about/research>.

[345] P Jones and P Elias, 'Administrative data as a research resource: A selected audit' (Report to the ESRC) (2006) 17.

[346] Department for Work and Pensions, 'Work and Pensions Longitudinal Study' <https://www.gov.uk/government/statistics/work-and-pensions-longitudinal-study>.

[347] Scottish Government, 'Linking social care, housing & health data' (2010) <https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2019/01/health-community-care-data-linkage/documents/linking-social-care-housing-and-health-data-social-care-clients-and-patients-views/linking-social-care-housing-and-health-data-social-care-clients-and-patients-views/govscot%3Adocument/Linking%2BSocial%2BCare%252C%2BHousing%2Band%2BHealth%2BData%2BSocial%2BCare%2BClients%2527%2Band%2BPatients%2527%2BViews.pdf>.

patterns of participation in the new deal for lone parents and assess the impact of the policy on the socio-economic status of claimants.[348]

The Office for National Statistics (ONS) has also been a pioneer in conducting administrative data research. In the late 1990s, the Office invested resources into developing a linked dataset including data from annual business surveys and data from the manufacturing sector to address research questions e.g. about productivity or growth about the characteristics of the UK market.[349] Similarly to the Department for Work and Pensions (DWP), the ONS hired and supervised within its premises in London academic or research institute experts with a two-fold aim: first, create and curate linked business datasets and, second, produce research outputs that will be published by the Office.[350] These pioneer attempts to use administrative data research for policy-making encouraged other departments in the UK to follow suit. The Ministry of Justice and HM Revenue and Customs collaborated with the DWP to carry out research on employment and benefit outcomes for ex-offenders,[351] whereas the Scottish and Welsh governments invested into developing innovative data linkage studies.[352] Such initiatives as the 'Fuel Poverty' in Wales aimed to link health and social care data (e.g. hospital admission records) with non-health data resources such as housing support data to assess the impact of the Welsh 'Warm Homes Nest' scheme on health outcomes.[353] In a similar vein, an interviewee working for the Department for the Economy in Northern Ireland confirms that their department would regularly 'commission research, primarily to be done by economists' on their administrative datasets.[354] What was, then, the impact of these pioneering efforts for the broader regulation of research data sharing in the UK?

[348] Smith et al (n 328) 16.

[349] F Ritchie, 'Business Data Linking – Recent UK experience' (2004) 33(1-2) Austrian Journal of Statistics 89.

[350] *Ibid* 91-92.

[351] Administrative Data Taskforce (n 23) 3.

[352] Scottish Government, 'Data linkage for research in Scotland' <https://www.gov.scot/publications/data-linkage-for-research-in-scotland/pages/scottish-government-funded-linkage-projects/>; Welsh Government, 'Fuel poverty data linking project: the impact on health of the Welsh Government Warm Homes Schemes' <https://gov.wales/fuel-poverty-data-linking-project-impact-health-welsh-government-warm-homes>.

[353] *Ibid* Welsh Government.

[354] Interview with participant nr. 16 (2018, phone).

As external users started producing valuable and tangible insights for policy-making, data custodians understood that had they been able to consistently engage experts from the academic sector with analysing their datasets, the likelihood of their investments paying off would be significantly bigger. To facilitate this, beyond commissioning research in the strict sense to address particular policy questions, departments started supporting 'infrastructure initiatives' that were deemed to be significant enough in terms of policy impact to warrant central government support.[355] Such initiatives were initially introduced in the health data context both in England and the devolved countries, e.g. the (English) Millennium Cohort Study,[356] the Scottish Longitudinal Study,[357] the Northern Ireland Longitudinal Study[358] and the SAIL Databank in Wales.[359]

The Department for Work and Pensions (DWP) was one of the first departments to proceed with more stable, infrastructural research data sharing initiatives involving external researchers. Within the 'New Deal for the Long Term Unemployed' project, Policy Studies Institute researchers enjoyed full access to DWP records, as well as were allowed to link such records with records from a survey conducted by the British Market Research Bureau International.[360] The HM Revenue and Customs, HM Treasury and the Department for Education facilitated similar initiatives such as the CASE/National Centre Micro-Panel Data Study and the Effective Provision of Pre-School Education study.[361] Does this mean that researchers started seeing in front of them a well-trodden and transparent path towards accessing administrative data?

---

[355] Smith et al (n 328) 5.

[356] *Ibid* 13: 'a major piece of ongoing research examining the health, family, and socio-economic context of children born in the UK in 2000-2001'.

[357] Scottish Longitudinal Study Development & Support Unit, 'Introduction to the Scottish Longitudinal Study' <https://sls.lscs.ac.uk/about/>.

[358] D O'Reilly et al, 'Administrative Data Research Northern Ireland (ADR NI)' (2020) International Journal of Data Population Science: 'NILS is an ongoing large-scale, representative data-linkage study created by linking data from the Health Card Registration system to the decennial Census returns starting at 1981 and to administrative data for example vital events such as births, deaths and marriages'.

[359] N Richards, 'Introducing the Administrative Data Research Centre Wales (ADRC-W)' (Blogs Wales Institute of Social & Economic Research, Data & Methods): 'themes such as education, health and social mobility data as well as many other long standing records from different government departments'.

[360] Smith et al (n 328) 16.

[361] *Ibid* 14.

Not necessarily. The significance of these projects notwithstanding, access to government data for social researchers was still exceptional and largely driven by the policy priorities of a data custodian and their subsequent appetite to commission or support research that would respond to government-designed research questions. An interviewee working for the Economic and Social Research Council, recalling their earlier experience within an administrative data research project, stresses that data sharing procedures were simplified due to the strong political buy-in from the Treasury that the project enjoyed, as well as the 'good personal relationships with the departments that would help unlock things'.[362] Two analysts working for the Welsh government concur that, at this historical time, data custodians largely relied on informal relationships and the circumstance of knowing someone within a particular government department, often due to the shared subject matter of interest between academics and policy-makers:

> *'Wales being a small country (...) people know their academics. There are teams within Knowledge and Analytical Services who are topic-specific as you might say. They get to know the academics within their field quite well and some would do commissioned work or analysis of linked data sets.'*[363]

Even though most commissioned research projects would be publicly open for competition under procurement rules, existing relationships between departments and academics shaped the scope of the offered research and, thus, the advantage of specific academics to earn the tender. As an interviewee who has worked for the Department for Education recalls, such relationships with particular academics would even result, exceptionally, in projects, particularly lower-cost ones, to be informally entrusted to such academics rather than being available for tender offers.[364] In these circumstances, most academic researchers were often ''put into a class of undesirables (…) labelled as 'intruders' (…)'',[365] with access to data being described as 'unreliable, haphazard and unregulated' from a research perspective.[366]

[362] Interview with participant nr. 19 (2018, Swindon).

[363] Interview with participants nr. 25 and 26 (2019, Cardiff).

[364] Interview with participant nr. 33 (2020, virtual).

[365] Jackson (n 308) 17.

[366] Elias (n 303) 184.

The fact that data custodians operated by reference to informal relationships unravels the complex processes of bargain and conflict within this phase. The gradual broadening of the space to make external use of government data for research purposes more stable presupposed that requesting researchers would need to negotiate the terms that custodians wanted to impose. With specific datasets coming at a premium, as they are uniquely held by specific departments, it is certain that such bargaining happened parallel to conflict and competition between multiple researcher groups which were interested in the same dataset or datasets. Inevitably, some researchers came out stronger from such conflicts, with the majority of academic social scientists being, at this time, not fully aware of the potential of government datasets and, thus, remaining 'at the margins of these developments'.[367] Academics who were interested in conducting their own, *independent* from government, research on government datasets would have to 'jump through many hoops' to be granted access, with government departments often relying on legal arguments to reject disclosure.[368] How was, then, independent social-science research on administrative datasets at all made possible?

iii. The origins of 'independent research': demanding independence, settling for informality

Being allowed to analyse government datasets not necessarily to respond to specific policy questions, but to pursue their independent research questions and interests, has been one of the consistent claims of researchers within this space. Far from an overnight development, however, the realisation of access to government data for independent social-science research was incremental and context-specific. In the absence of a centralised statistical authority within the UK that would provide a more 'regulated' approach,[369] each department had to develop its own data sharing policy in response to requests for various purposes, including research. In the late 1990s, the previously mentioned government initiatives started to yield published

[367] Smith et al (n 328) 5.

[368] Ritchie (n 342) 104.

[369] Elias (n 303) 186.

outcomes and this stimulated academic interest in using administrative data for social-science research.[370]

The number of requests gradually increased as the research potential of administrative data became clearer to researchers. An interviewee who has worked for the Department for Education concedes that they were 'quite surprised by the scale of requests' coming from academics, compared to what they were used in seeing at their previous position at the Department for Work and Pensions.[371] Compared to previously used data sources like official statistics collated and published by statistical agencies,[372] administrative datasets provided a more comprehensive option for understanding social and economic patterns.

Researchers' interest prompted high-level standard-setters and data custodians to provide more clarity on the possibilities of research data sharing in the UK. With the central government commissioning as part of the UK National Data Strategy a report on the available public-sector data assets for potential research use,[373] specific data custodians like the Office for National Statistics (ONS), excited about the potential benefits for policy-making, started encouraging academics to work on their datasets.[374] An interviewee who has worked for the ONS provides some clarity of the Office's efforts to formalise access to its datasets for independent, external researchers. As they explain, the office had already been facilitating access for interested researchers since the 1990s, using a legally controversial method, through which external researchers were nominally appointed by the Office in the interest of allowing them to conduct research on its administrative data. They eloquently recall how things moved from this state of affairs to formalising data access in the early 2000s:

> *'When I joined ONS, the person who employed me, the then Chief Economist at ONS (…) said, we've got this data which is a bit more sensitive, it's all business data, we got a couple of people coming in from different organisations and they physically use our computers or we let them sit on our own computers. It's not*

[370] E.g. J Ball and M Marland, 'Male Earnings Mobility in the Lifetime Labour Market Database' (1996) (Analytical Services Division, Department of Social Security Working Paper Series).

[371] Interview with participant nr. 30 (2020, virtual).

[372] Jackson (n 308) 18.

[373] Elias (n 303) 186.

[374] Ritchie (n 71) 30.

> *a very sustainable model, and we are also not sure about the legal basis for this. Essentially, can you turn this around to something (…) more than an ad-hoc arrangement?* '[375]

The same interviewee suggests that a lot of other departments were using similar mechanisms to provide data access to external researchers 'within the spirit of working for the public good'.[376] Increasing academic and policy interest, however, does not necessarily mean that data access for research would be automatically streamlined across the UK public sector. In most cases, procedures for research data sharing did not exist and academic requests often stumbled upon legal concerns about the status of administrative data and technical difficulties with regard to preparing such data for research use.[377]

Bargaining between government departments and academics, within the already existing contractual relationships between them, was crucial in overcoming these barriers. Some research groups managed to gain consistent access to specific datasets for the purposes of producing their own independent research.[378] As part of the agreement, researchers would, in turn, commit to disseminate the results of their work to improve knowledge of the general public, as well as feed back to the government analysts with a view to strengthening the existing evidence-basis for evaluating government policy.

This idea of 'partnership' between the academic research community and the government was embraced by key data custodians in England.[379] The Department for Work and Pensions (DWP) granted academics access to use some of its datasets to study issues relating to labour market transitions and earnings,[380] particularly when requests for projects corresponded to its key 'areas of research interest'.[381] Significant administrative data research projects were conducted in this period. Elias, an established administrative data researcher, led a couple of these projects: he used Higher Education Statistics Agency data to create a longitudinal sample

[375] Interview with participant nr. 42 (2020, virtual).

[376] *Ibid.*

[377] Ritchie (n 71) 29.

[378] Smith et al (n 328) 5.

[379] Jackson (n 308) 25.

[380] Jones and Elias (n 338) 21.

[381] Department for Work and Pensions (n 337).

of graduate students to track their entry into the labour market, as well as the DWP's JUVOS database to assess the impact of the Employment Zones programme on unemployed individuals.[382] Hales, another leading researcher, and his team contributed to the evaluation of the Employment Zones programme by using another DWP dataset, JSA, to identify potentially eligible participants who had failed to make use of the programme to get out of unemployment.[383] DWP data was also used by other academics for projects focused on analysing individual level claimant dynamics in the context of housing benefits.[384] Beyond the DWP, not many partnerships between academics and government departments are recorded from this period, although the Department for Education is also reported to have facilitated access to its National Pupil Database for researchers to work on pupil microdata.[385]

The language of 'partnership', also popular within public management drives for public-private partnerships in general,[386] naturally, obfuscates power imbalances in custodian-researchers' relationships. In the previously mentioned cases, there was no established and publicly documented process or 'access route' for interested academic researchers.[387] Like commissioned research, academics who worked in a particular research area would be better positioned in making contact with data custodians in an 'ad personam way, rather than a necessarily open or transparent way'.[388] As a result, the availability of certain datasets was communicated to a 'limited circle of researchers' who 'knew the correct people to speak to' within the relevant departments.[389] Processes of conflict and bargaining, between researchers and custodians and between researchers and other researchers, shine through their absence when seeing these 'partnerships' from some distance.

[382] Smith et al (n 328) 19.

[383] J Hales et al, 'Evaluation of Employment Zones: Report on a Cohort Study of Long-Term Unemployed people in the Zones and a Matched Set of Comparison Areas' (2003) (National Centre for Social Research).

[384] M White, 'Macro evaluation of the New Deal for Young People' (2000) (PSI briefing note).

[385] Jones and Elias (n 338) 4.

[386] McLaughlin et al (n 333).

[387] A Sexton et al, 'A balance of trust in the use of government administrative data' (2017) 17(4) Archival Science 305.

[388] Interview with participant nr. 20 (2018, phone).

[389] Sexton et al (n 380) 313.

Coming closer to the intricacies of such partnerships, however, one sees a different picture. An interviewee who has worked for the Department for Education concedes that data disclosures at this time 'wouldn't be to some random individual – it would only be with someone we've worked with before'.[390] Administrative data researchers working in Scotland corroborate this picture, reflecting on how their personal relationships with individuals working for the data custodians enabled them to get access:

> *'I mentioned the informal there and I think that actually was crucial in moving forward the work (…) through individuals within National Records of Scotland, we were really able to start moving forward, thinking how we would enable the data to be made available for linkage and later for analysis.'*[391]

> *'My PI had actually negotiated access beforehand. I think it was an easier task for him, because he was part of the research team that conducted the PSE survey (for the DWP).'*[392]

> *'Because of the funding, I already got to know them from the very first day of my PhD. They had the best available data. Not brilliant, but the best available for national linkage. (…) Other people had to change their PhD basically. I suppose because of the funding and the will to share this data, this was more feasible.'*[393]

As a result, access to government data for social-science research in the late 1990s and early 2000s was far from transparent, predictable and equally feasible for less well-connected academics, and more of a 'patchy landscape' with researchers being unable to predict whether and when they would be granted access.[394] Nonetheless, studies like the ones mentioned in this section, which were made possible on the basis of informal and ad-hoc arrangements, yielded significant policy benefits, elucidating the socio-economic impact of particular policies on

[390] Interview with participant nr. 33 (2020, virtual).

[391] Interview with participant nr. 34 (2020, virtual).

[392] Interview with participant nr. 38 (2020, virtual).

[393] Interview with participant nr. 37 (2020, virtual).

[394] Interview with participant nr. 21 (2019, Swansea).

target populations and throwing light on questions which were previously obscured by high non-response rates to government surveys.[395] This fuelled the appetite of high-level standard-setters such as the central government to explore how academic data access for social-science research could be consistently widened and formalised.

iv. Widening research data access: the view from the high level

In 2002, the Cabinet Office published a set of recommendations about making use of government data more 'widespread and coordinated'.[396] These recommendations reflected an increasingly held appreciation that administrative data are a 'largely untapped source of information for research and policy evaluation in the UK', the potential of which can only be fully realised if departments become more transparent about their data sharing policies.[397] In the light of statistical research methodologies also becoming more sophisticated in the early 2000s, central government encouraged public-sector data custodians to consider facilitating 'general research not directed by the department' in a more streamlined and systematic way.[398] An interviewee who has worked for the HM Revenue and Customs (HMRC) provides some background on their department's motivation to set a 'uniform data sharing strategy':

> *'So there were two reasons driving towards the data sharing strategy - one was (…) a recognised risk that HMRC wasn't making the best use of the business intelligence it had - that translates as it was not exploiting data as best or as effectively as it could do. (…) And then there is a stated objective from the HMRC strategy that HMRC will use its assets across government for public benefit or benefit of other departments.'*[399]

This was, certainly, not only in the minds of HMRC officials. The central government was generally keen on seeing more policy-relevant social-science research on government data in such areas as evaluating the long-term impact of drugs or health policies, school performance,

[395] Smith et al (n 328) 13.

[396] Jones and Elias (n 338) 69.

[397] Administrative Data Taskforce (n 23) ii.

[398] F Ritchie and R Welpton, 'Sharing risks, sharing benefits: Data as a public good' (Conference Paper) (2011 Eurostat) 4.

[399] Interview with participant nr. 32 (2020, Oxford).

the connection between re-offending and unemployment and the connection between tax cuts and social benefit provision.[400] The research data sharing arrangements of the time, i.e. informal and ad-hoc procedures governing access based on pre-existing relationships between a small number of academics and public officials, were hardly satisfactory in that regard. Access to government data for social-science research purposes had to become more predictable, organised and reliable, in the interest of attracting the interest of expert researchers that can, at a low cost for the government, contribute to the production of a robust evidence-basis for UK policy-making. This was also in line with shifts to more technocratic methods of managing the civil service to boost efficiency, prevalent within the new public management approach.[401] What were, however, the major factors driving the production of such an evidence-basis by the actors within the research data sharing space?

B. Formalisation: Creating Formal Frameworks to Streamline Research Data Sharing in the UK

The process of formalising access to administrative data for social-science research purposes was motivated by two major narratives, co-constructed and employed by not only central government and data custodians, but also academic researchers and funding bodies or research consortia. I elaborate on these two narratives, 'efficiency' and 'accountability', before presenting the first examples of formalised research data sharing frameworks in the UK public sector. Then, I discuss the major effort to expand formalisation across the public sector furthered by the Administrative Data Research Network investment, as well as the resulting crystallisation of specific stages of research data access. The following section (C) will interrogate whether this crystallisation corresponds to the adoption of transparent and predictable data sharing frameworks by data custodians across the UK public-sector *in practice.*

i. 'Independent' Research Data Access, 'Efficiency' and 'Accountability'

While improving collective knowledge about social and economic behaviour can be seen as an end-in-itself, coordinating individual and organisational behaviour to streamline data access

[400] Administrative Data Taskforce (n 23) 1.

[401] McLaughlin et al (n 333).

for social-science research purposes required commitment to some greater underlying aim. Why should government departments be carrying out this function at the risk of incurring logistical and reputational costs?[402] Two major narratives emerged to promote the production of better knowledge through streamlining research data access and, thus, facilitating more administrative data research in this space: achieving higher levels of government 'efficiency' and strengthening government 'accountability' to its citizens and organised political stakeholders. Although both relate to the operation of government, these narratives have been strategically employed, as it will be shown, by various actors populating the research data sharing space, e.g. researchers, funding bodies and data custodians.

The idea of 'efficiency' hints at both strictly construed financial gains in delivering public services and the broader aim of policy improvement. In the advent of long-term austerity measures and constant grappling with finding adequate resources for public service, many departments aimed to harness the potential of administrative data research to inform resource allocation.[403] The potential of administrative data research to yield financial gains was not only adopted by departments with remit to produce statistical research in the first place,[404] but more generally by departments keen to maximise 'the returns on the resources they invest in data collection by exploiting administrative data to the full'.[405]

Crucially, such departments saw the formalisation of research data access as a necessary step to stabilise researcher expectations and thus guarantee consistent analytical interest in their data in the long term. These departments did not see research data sharing as a one-off activity, but invested into the 're-use' of their data by external researchers.[406] An interviewee who has worked for the HM Revenue and Customs highlights better debt management and the reduction of debt as one of the key incentives behind the establishment of data sharing procedures in their department.[407] Another interviewee working for the Office for National Statistics clarifies that one of the primary purposes behind sharing data to external researchers was the aim of

[402] Ritchie (n 342) 105.

[403] Scottish Government (n 345).

[404] Ritchie and Welpton (n 324) 4.

[405] R Welpton and M Wright, 'HMRC Datalab: Engaging with the External Research Community' (2011) (ESRC) 1.

[406] Administrative Data Taskforce (n 23) ii.

[407] Interview with participant nr. 32 (2020, Oxford).

'measuring the effects of policies and ensure that (…) government (…) spends money in a more effective way'.[408] The fact that academics are willing to conduct such research 'for free' made this type of data sharing even more appealing for departments.[409]

Beyond financial or 'monetary' gains,[410] departments aspired to become more 'efficient' in tackling complex policy issues, the root causes of which had hitherto not been well-researched. Both in respect of challenging or triangulating existing knowledge, research and statistics produced by the government,[411] increased access to administrative data promised to enable decision-makers to gain novel insight into multi-layered policy questions by linking large scale administrative datasets belonging to various data custodians.[412] For example, the Welsh Government has sought to explore the relationship between social care, health and housing support within their 'Fuel Poverty' project.[413] More 'efficient' and accurate interventions on behalf of the government can, the narrative goes, stimulate economic growth and improve the health and well-being of the British people.[414] Within the 'Fuel Poverty' data linkage project, the Welsh government measured the extent to which a fuel benefit policy actually led into an improvement of the health of vulnerable Welsh citizens.[415] One of this study's interviewees, who worked in this project, stresses that making clear the 'big evidence gap' that existed and

[408] Interview with participant nr. 2 (2017, Titchfield). Note that the ONS, due to its organisational mission to produce high-quality statistical research that scrutinises policy-making and implementation in the public sector, is an idiosyncratic case and the present quote shall not be seen as indicating the existence of a broader (meta)narrative, i.e. 'effectiveness', that is not covered by the two identified narratives of 'efficiency' and 'accountability'.

[409] Interview with participant nr. 30 (2020, virtual).

[410] Interview with participant nr. 21 (2019, Swansea).

[411] C Allsopp, 'Review of Statistics for Economic Policymaking: Final Report to the Chancellor of the Exchequer, the Governor of the Bank of England and the National Statistician' (2003) 40.

[412] Jones and Elias (n 338) 76.

[413] Scottish Government, 'Linking social care, housing support & health data for statistics: Project Plan and Privacy Impact Assessment (PIA)' <https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2019/01/health-community-care-data-linkage/documents/linking-social-care-housing-support--health-data---privacy-impact-assessment/linking-social-care-housing-support--health-data---privacy-impact-assessment/govscot%3Adocument/Linking%2Bsocial%2Bcare%252C%2Bhousing%2Bsupport%2B%2526%2Bhealth%2Bdata%2B-%2BPrivacy%2BImpact%2BAssessment.pdf>.

[414] Administrative Data Taskforce (n 23) 2.

[415] Welsh Government (n 345)

the 'health impact' that their project would have was crucial to establish and implement a process for data sharing.[416] With evidence-based policymaking increasingly regarded as an 'essential part of good governance',[417] commentators have gone as far as suggesting that policy improvement is a prerequisite for the legality of granting researchers access to administrative data.[418]

On the other hand, the notion of 'accountability' draws on ideals of transparency and openness in governmental decision-making, with a view to consolidating public trust in the executive.[419] If data custodians realised the 'efficiency' of research data sharing after long processes of bargaining with researchers and intermediaries, 'accountability' ideals have been generated by decades of political conflict and bargaining about the transparency that the central government owes to the citizens about its decision-making. The UK government has repeatedly committed over the last few decades to 'empowering citizens' with information that will allow them to 'challenge government if they see inefficiencies'.[420] Independent' research on government data, i.e. research conducted by a neutral individual or group, ideally when the latter subscribe to recognised standards of professional and ethical integrity as in the case of academic researchers, has been regarded as integral to realising this commitment.[421] This aspiration stemmed from a broader policy agenda on 'greater openness' as part of a 'global movement towards transparency' and 'open policy making',[422] premised on the idea that

[416] Interview with participant nr. 25 (2019, Cardiff).

[417] HM Government, 'Open Data White Paper: Unleashing the Potential' (June 2012) 41: 'It is increasingly clear that evidence-based research and policy are fundamental to understanding and responding effectively to the challenge of assuring the health and social and economic wellbeing of people'.

[418] Jackson (n 308) 22: 'It follows that the necessity test is met best when a research project is complementary to the analysis of the department. Interference with the privacy of the department's clients may be necessary if the outcome is evidence that fills a knowledge gap in the department or provides elaboration or greater context and insight to the evidence the department already has'.

[419] HM Government (n 410) 5: 'Data is the 21st century's new raw material. Its value is in holding governments to account'.

[420] *Ibid* 12.

[421] M Woolard, 'Administrative Data: Problems and Benefits. A Perspective from the United Kingdom' in A Dusa et al (eds), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences* (Scivero 2014) 49.

[422] HM Government (n 410) 5; HM Government, 'Open Government' <https://www.gov.uk/government/collections/open-government>.

factual evidence underpinning policy decisions shall be made public to allow citizens 'scrutinise the work of those they elect'.[423] Hence, by streamlining access to government data for independent social-science research, the government aspired to realise the value of data 'in holding governments to account (…) (and) driving choice and improvements in public services'.[424]

An interviewee working for the Office for National Statistics argues that there is 'public benefit' in external researchers using the Office's data to 'prove that our statistics are incorrect' or to 'produce a better measure of inflation', even if there is no 'direct use' of these findings for public policy.[425] Another interviewee working for the data custodian body of Honest Broker Service in Northern Ireland draws a connection between the idea of accountability and the professional identity of 'statisticians', shared by many individuals working for data custodians:

> *'I don't know whether this is simply a Northern Irish Statistics and Research Agency culture, we are statisticians, we like things to be open, completely clear and transparent to everybody. This is part of our independence as statisticians (…) putting all the information in the public domain about what research is this, when it is happening, what it is about, when the application came in etc.*'[426]

Research output, produced from research on administrative data, is to be treated as a 'public good' under this narrative.[427]As one of the interviewees mentions, this is also reflected in the Office for National Statistics (ONS) requirements that all outputs from administrative data research are published on open-access websites because:

> *'A lot of the projects evaluate Government's policies, evaluate government delivery. That is very important for democratic debate, for the public to actually*

[423] Department for Digital, Culture, Media & Sport, 'UK National Action Plan for Open Government 2019-2021' (2019).

[424] HM Government (n 410) 5.

[425] Interview with participant nr. 1 (2017, Titchfield).

[426] Interview with participant nr. 12 (2018, Belfast).

[427] Ritchie and Welpton (n 391).

> *know how effective the government is. Ensuring that transparency and accountability exist.'*[428]

An interviewee who has worked for the ONS argues that the Statistics and Research Service Act 2007 rebutted the previous presumption that data is not available for research unless the custodians provide them for this purpose. Essentially, this is to make a bold claim all data are 'fair game' for researchers unless custodians can justify withholding them.[429] Another interviewee working as an administrative data researcher in Northern Ireland expresses the same ideal in the following words:

> '*There has to be an aspiration (…) that this is public data and (…) if you have an ethically approved, methodologically sound project that addresses something that has public benefits, then we should be advocating for this project to be supported.'*[430]

Following this line of thought, various organisations have explicitly committed to realise the potential of 'independent' administrative data research to make government policy-making not only more efficient, but also more accountable. One of the first research data infrastructures in the UK public sector to facilitate streamlined access of external researchers, the HM Revenue and Customs (HMRC) DataLab, was established in fulfilment of HMRC's commitment to the government's 'Transparency Agenda'.[431] The benefits of transparency were associated by the HMRC with both the aim of 'promoting social and economic growth' and that of promoting 'public understanding of its work, (…) accountability and public confidence'.[432]

More recently, in the context of digitalising the UK justice system, the CEO of HM Courts and Tribunals Service (HMCTS) committed to making data, 'in a suitably anonymised way',

[428] Interview with participant nr. 3 (2017, Titchfield).

[429] Interview with participant nr. 42 (2020, virtual).

[430] Interview with participant nr. 10 (2018, Belfast).

[431] D Bega, 'The HMRC Datalab: Sharing administrative and survey data on taxation with the research community' in A Dusa et al (eds) (n 414) 61.

[432] *Ibid* 64.

available for researchers and academics to use.[433] This commitment was elaborated upon in later HMCTS publications, where it was clarified that 'independent research' on the reform programme is integral to establishing accountability and transparency.[434] Both HMRC and HMCTS have acknowledged the need to improve the way they share data with external researchers, who are facing a number of difficulties in accessing their data.[435] Such researcher organisations as the Education Policy Institute have also pledged their commitment to use administrative data research to 'hold government (…) to account for their performance' and help improve policy-making.[436] These aspirations materialised through both public demands and bargaining in the creation of the first formalised data sharing frameworks and infrastructures for external researchers in the UK public sector.

ii. Formalising 'Independent' Research Data Access: Trailblazers and the Ambition for Wider Formalisation

Like its role in pioneering government-led administrative data research, the Office for National Statistics (ONS) established the first data infrastructure for independent research on government data in 2004, the Virtual Microdata Laboratory (VML).[437] Driven by the central government's interest to better understand the productivity of UK manufacturing,[438] the VML was launched to provide a 'secure facility for access to confidential business micro data across all ONS sites'.[439] In its first few years of operation, the VML significantly streamlined external researcher access to ONS data, starting with supporting six (6) research projects in 2004 and reaching the number of eighty-nine (89) projects by 2008 and more than two-hundred thirty

[433] HMCTS Chief Executive, 'Modernising the Courts and Tribunals Service: Future of Justice Conference' (14 May 2018) <https://www.ucl. ac.uk/laws/sites/laws/files/ucl_foj_01_03_acland-hood.pdf>.

[434] Ministry of Justice, 'Evaluating our reforms: Response to PAC Recommendation 4, January 2019' <https://www.gov.uk/government/news/moj-response-to-public-accounts-committee-transforming-courts-and-tribunals>.

[435] N Byrom, 'Digital Justice: HMCTS data strategy and delivering access to justice Report and recommendations' (October 2019) <https://research.thelegaleducationfoundation.org/blog/digital-justice-hmcts-data-strategy-and-delivering-access-to-justice> 31; M Almunia et al, 'Expanding access to administrative data: the case of tax authorities in Finland and the UK' (2019) 26 Int Tax Public Finance 661-676.

[436] Education Policy Institute, 'About Us' <https://epi.org.uk/our-mission-and-purpose/>.

[437] Jones and Elias (n 338) 39.

[438] Welpton and Wright (n 398) 11.

[439] Jones and Elias (n 338) 39.

(230) by 2017.[440] Research outputs on such issues as the impact of innovation to productivity or the calculation of labour cost assessments for the National Health Service were published and 'influenced policy-making at all levels'.[441]

Such stellar research outputs were instrumentalised by the research community in their efforts to convince the ONS to make data access even more transparent and predictable. Indeed, beyond the particular research projects, the VML has gradually come to be widely regarded as a 'model of best practice (…) emulated across the UK and world-wide'[442] and a 'major resource for academic researchers'[443] due to its establishment of fundamental governance principles for external researcher data access to government data. The ONS has published comprehensive guidance on the its decision-making procedure and interpretation of its governance requirements, as well as a four-week timescale within which researchers will normally receive a response to their request.[444] More specifically, the ONS created a specific governance infrastructure, the Micro-Data Release Panel (MRP), and entrusted it with responsibilities to manage the strategic and operational aspects of research data access to the VML. External researchers, interested in using ONS datasets, have to submit an application to the MRP, convincing the panel that their request is 'related to a viable and relevant research topic' and that they themselves are able to undertake high-quality research that will 'add significant value to existing research in that area'.[445]

Crucially, the ONS has not only published guidance, but, in fact, consulted the research community about the appropriateness of its criteria in July 2016. As an interviewee working for the Office explains, the data custodians themselves were not satisfied with the previous criteria because they were 'a lot vaguer and gave us a lot more room for manoeuvre', aiming instead for a 'set of criteria that could then be applied consistently throughout'.[446] Another

[440] Ritchie (n 71) 31; interview with participant nr. 1 (2017, Titchfield).

[441] Ritchie ibid 29; F Ritchie, 'Access to business microdata in the UK: dealing with the irreducible risk' in Eurostat, *Monographs of official statistics: Work session on statistical data confidentiality* (2005) 239.

[442] Welpton and Wright (n 398) 12.

[443] Ritchie (n 71) 29.

[444] *Ibid*.

[445] Jones and Elias (n 338) 51.

[446] Interview with participant nr. 1 (2017, Titchfield).

interviewee working for the ONS highlights the significance of formalisation for external researchers by referring to the function of access decisions as precedents for other researchers:

> *'Over the years, we built up a system of precedent decisions (…) we decided we had to rework the whole process and formalise it a lot more. (…) So now we have a written set of precedents, which allow us to make decisions and the understanding is this is a lot more formal and transparent procedure (…) We want to make sure that everybody understands the precedents and we apply them consistently and correctly.'*[447]

The Office for National Statistics (ONS) is, then, a clear example of widening the scope of appropriate access to administrative data for research purposes through collaborative modes of dialogue and bargaining through invoking tangible outcomes. This example inspired the Economic and Social Research Council (ESRC) to invest resources into the creation of a similar to the Virtual Microdata Laboratory (VML) research data infrastructure to facilitate access to broader public-sector data assets. In 2009, the Secure Data Service (SDS) was launched within the ESRC-supported UK Data Archive, allowing for the first time in the UK remote access for academic researchers via their accredited institution's 'safe room'.[448] As an interviewee who has worked for both the ONS and the SDS recalls, the latter's data sharing policy was largely influenced by the need to demonstrate conformity with the VML framework, in the interest of enhancing data sharing collaboration and technical compatibility between the two infrastructures.[449] The same interviewee highlights the significance of remote access for attempts to formalise access for the majority of the external research community, noting that previous models like the VML unavoidably privileged researchers who happened to live closer to their secure research environments ('safe settings').[450]

Nonetheless, infrastructures are not used unless more data custodians have the appetite to grant access to their datasets. In 2011, HM Revenue and Customs (HMRC) were one of the first major ministerial departments in the UK to formalise access to their datasets for academic

[447] Interview with participant nr. 2 (2017, Titchfield).

[448] Welpton and Wright (n 398) 12.

[449] Interview with participant nr. 28 (2020, London).

[450] *Ibid.*

researchers 'working independently' through the 'HMRC DataLab'.[451] Only allowing access on-site, HMRC sought to establish a secure research environment within which researchers would access 'anonymised taxpayer and customs data for research and analysis purposes',[452] with a view to producing high-quality research that would benefit both the HMRC and the academic community. Mirroring the Office for National Statistics governance procedures, HMRC created their own Microdata Release Panel. This Panel handles operational aspects of data sharing, reviewing research proposals and performing output validation and disclosure control checks to assess identifiability risks.[453] Researchers applying for access to the HMRC DataLab have to demonstrate that their proposed research will 'deliver a public benefit to society'.[454] As an interviewee who has worked for the HMRC explains, the latter would be interpreted by reference to HMRC functions, covering such areas as 'understanding compliance, taxation or commercial activity'.[455] This demonstrates that researchers' bargaining does not always have the same results, with HMRC adopting a narrower criterion to govern data access compared to the Office for National Statistics.

Formalisation gradually expanded to other public bodies, in the interest of attracting more researchers by reassuring them that they will no longer 'have to know the right person' to access data in a predictable timescale.[456] For example, the Department for Education (DfE) established its own Data Sharing Approval Panel and approval process, providing guidance to researchers interested in accessing such datasets as the National Pupil Database and the Individualised Learner Record.[457] An interviewee who has worked for the DfE recalls that the Panel would assess whether the requester is a trusted organisation and the data will be used appropriately, often engaging in communication with them to clarify the scope of the data use.[458] Furthermore, the DfE funds an National Pupil Database User Group, hosted at the

[451] Bega (n 424) 68.

[452] *Ibid* 66.

[453] *Ibid* 66.

[454] Welpton and Wright (n 398) 7.

[455] Interview with participant nr. 32 (2020, Oxford).

[456] Sexton et al (n 380) 313.

[457] HM Government, 'How to access Department for Education (DfE) data extracts' (2018) <https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts>.

[458] Interview with participant nr. 33 (2020, virtual).

University of Bristol, to encourage better communication between academics using the Database and communicate to them the availability of new data extracts.[459]

Beyond England, the Honest Broker Service (HBS), part of the Business Services Organisation within the Health and Social Care department of the Northern Irish government, is one of the luminary data custodians in formalising external researcher access to its resources.[460] The HBS created 'safe havens' for secure data linkage and de-identification, allowing access to anonymised data for researchers interested in the social dimensions of health and social care.[461] An interviewee working for the HBS confirm that they have established their own Governance Board and follow a similar decision-making procedure with other departments:

> *'So, we have our discussion, an in-depth discussion, about what their intentions are, what the aims of the project are, perhaps tease out a bit more the variables they are interested in, (…) mindful of the requirements of the Governance Board, what they will be looking to see in that application, and there are some sets of criteria that it must meet.'*[462]

How widespread were, however, such initiatives across the UK public sector? In spite of this gradual expansion, the truth is that the vast majority of the UK public-sector data custodians remained somewhat cryptic and inconsistent as to the ways in which they handled access to their data for independent research purposes. To encourage broader consistency and transparency through widespread formalisation of research data sharing, the central government, through the Economic and Social Research Council, funded in 2011 the establishment of the Administrative Data Taskforce (ADT).[463] The taskforce brought together all the major actors of this space: senior academic researchers, representatives from funding

[459] Centre for Market and Public Organisation, 'NPD User Group' <http://www.bristol.ac.uk/cmpo/npd-user-group/>.

[460] G Robinson et al, 'Public Attitudes to Data Sharing in Northern Ireland: Findings from the 2015 Northern Ireland Life and Times survey' (2018) (ADRC-Northern Ireland) 4.

[461] *Ibid.*

[462] Interview with participant nr. 14 (2018, Belfast).

[463]Administrative Data Taskforce (n 23); interview with participant nr. 19 (2018, Swindon).

bodies, data custodians, as well as government policy-makers.[464] ADT's organisational mission was to provide recommendations for widening the use of administrative data by independent researchers, with a view to creating a 'robust UK-wide evidence base' that would both improve policy making and inform 'evaluation of government policy' by the academic community.[465]

In this direction, the taskforce recommended the establishment of a UK-wide Administrative Data Research Network (ADRN), accompanied by one Administrative Data Research Centre in each of the four UK countries.[466] Through this network, the taskforce suggested, researcher access to administrative data would become much more uniform and streamlined. This would be assisted by establishing common training and accreditation processes, as well as common approval requirements. Crucially, especially for the latter aim, the central government relied on the collaboration between departments (or other data custodians) and the UK research community.[467] How did this collaboration work out in practice?

iii. The ADRN (2013-18) as an Agent of Systematic Formalisation

Since its establishment and funding by the Economic and Social Research Council (ESRC) in 2013, the ADRN worked towards extending the formalisation of research data sharing processes from a handful of pioneer data custodians to the vast number of public-sector bodies in the UK.[468] In the words of an interviewee working for the ESRC, the aim of the Network was to use 'the wealth of the UK's administrative data for the purpose of scientific research, in social science and economics'.[469] With 'independent' – understood as 'original' and 'non-commissioned'[470] – research at the heart of its mandate, the network established a clear and transparent process to handle data sharing requests, assisting academics in preparing their applications. In that sense, the ADRN was the first intermediary body to go beyond general advice and guidance to personalised support in this space. The ADRN's Approvals Panel would

[464] Elias (n 303) 188.

[465] Administrative Data Taskforce (n 23) ii.

[466] Woolard (n 414) 50.

[467] Administrative Data Taskforce (n 23) ii.

[468] Elias (n 303) 185.

[469] Interview with participant nr. 19 (2018, Swindon).

[470] Administrative Data Taskforce (n 23) 6.

take into account a set of key considerations: the feasibility of the proposed study, its ethical acceptability (considering the lack of consent), its scientific merit and potential for yielding benefits to the public, as well as the risks from a privacy and data protection perspective.[471] Applicants, supported by ADRN officers, would submit their research proposals, often receiving feedback and required amendments by the Approvals Panel.

The network aspired to promote these criteria as core elements of a research data sharing process across the public sector through its commissioning panel (the 'ADRN Board').[472] The panel, which was chaired by a senior academic, comprised of government statisticians and ESRC representatives, as well as of representatives from major data custodians such as the DWP, the HMRC and the governments of the devolved countries.[473] The existence of the four Administrative Data Research Centres in the UK countries served both as a reflection of the devolved nature of government functions in the UK and as a way of developing consistent data sharing standards across different jurisdictions.[474]

Through this collaboration between academics and government departments in the four countries, the Network aspired to 'leverage' the expertise of its members and supported researchers to trigger wider formalisation of researcher data access and streamline the production of independent administrative data research.[475] Viewed from the perspective of regulatory space, this might be seen as a more adversarial and blunt approach compared to previous organisations in this space that might have tried to negotiate wider access to administrative data through more incremental steps. As the following section will explain, this effort might have been overambitious in respect of maximising actual data acquisition from

[471] UK Statistics Authority, 'First Meeting of the Administrative Data Research Network Board' (2014) <https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-minutesandpapersfromadrnboardon14april201_tcm97-43879.pdf>.

[472] UK Statistics Authority, 'Meet the ADRN Board' <https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/adrn/meet-the-adrn-board/>.

[473] Elias (n 303) 190.

[474] *Ibid* 185.

[475] Interview with participant nr. 19 (2018, Swindon).

government departments. The Network faced barriers that made access to Government-held administrative data very difficult and, at times, impossible.[476]

Does this mean that the Administrative Data Research Network (ADRN) was a failure? Regardless of actual data access, the Network was a resounding success in terms of making the demand for wider formalisation of research data sharing much more mainstream in the UK public sector. By sponsoring more than one-hundred and twenty (120) independent research projects by academics and negotiating access with a vast number of data providers across the UK,[477] the ADRN contributed significantly to the crystallisation of specific research data access procedures. These procedures and their accompanying guidance, covering legal, ethical and information security aspects of data sharing, are now recognised as best practice and adopted by many data custodians in the UK.[478] Broadly formulated, these aspects relate to four 'stages' of data sharing: de-identification of the data to mitigate disclosure risks from a data protection and privacy perspective,[479] public interest and ethical review of the research proposal, accreditation and training of the individuals that will be accessing the data and data security of the working environment within which the research will be conducted.[480] In developing this data sharing framework, the Network largely drew on pre-existing practice in the context of health data sharing for research and the '5 safes' framework developed by the Office for National Statistics.[481]

[476] In its fourth year of operation, the Network had managed to get the data for circa 19% of the ADRN-approved projects it had been facilitating, see UKSA (n 15) 14.

[477] *Ibid*.

[478] D Hand, 'The Role of the ADRN Governing Board' (2013) <https://esrc.ukri.org/files/research/administrative-data-taskforce-adt/the-role-of-the-adrn-governing-board-david-hand/>.

[479] On the meaning of de-identification as 'a process which reduces the risk of direct identification, but which does not produce anonymous data', see M Mourby et al, ''Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK'' (2018) 34(2) Computer Law & Security Review 222, 224.

[480] Jackson (n 308) 22.

[481] C Brett and I Deary, 'Realising health data linkage from a researcher's perspective: Following up the 6-Day Sample of the Scottish Mental Survey 1947' (2014) 5(3) Longitudinal Life Course Studies 283; Desai, Welpton and Ritchie (n 322) 5: 'safe people, safe projects, safe data, safe settings, safe outputs'.

By simplifying the 'complex discussion around data access into a set of related but independent questions',[482] the ADRN facilitated the adoption of consistent and formalised research data sharing standards by various data custodians in the UK.[483] These standards have been endorsed by the Information Commissioner's Office (ICO)[484] and were recently enacted, to a significant extent according to some interviewees who have worked for the Office for National Statistics,[485] into formal UK legislation. Sections 64 and 70 of the Digital Economy Act 2017 empower all public authorities to disclose data for research in accordance with operational requirements and accreditation criteria published by the UK Statistics Authority,[486] largely mirroring the standards adopted by the ADRN.

In the light of these developments, the former ADRN Director has claimed that administrative data researchers in the UK 'can plan their research and be confident that access to administrative data will be granted'.[487] Does this claim, however, reflect adequately the realities of implementing research data sharing frameworks in the UK public sector today? A discussion of the following, and last, historical phase within which actors shaped the development of research data sharing in the UK indicates the existence of significant limitations to the formalisation of this activity across the public sector, with ideals of 'accountability' often giving rise to 'efficiency'-driven priorities.

C. Fragmentation: Unequal Access of Researchers to Administrative Data in the UK

Contemporary research data sharing realities in the UK indicate that formalisation is rather a work in progress than a widely applied reality. With 'accountability' reasons for consistent and predictable data access for independent researchers gradually being eclipsed by 'efficiency' rationales, a fragmented and unequal picture arises as an outcome of the processes of conflict and bargaining between the actors of this space. I sketch the contours of this fragmentation,

---

[482] *Ibid.*

[483] E.g. O'Reilly et al (n 351) on the Northern Irish Statistics and Research Agency: 'NISRA hosts the Secure Environment where accredited researchers access data for ADR NI research projects. This operates under the ''Five Safes'' principles which have been endorsed by the Information Commissioner's Office'.

[484] *Ibid.*

[485] Interviews with participants nrs. 1 and 4 (2017, Titchfield), 42 (2020, virtual).

[486] UK Statistics Authority (n 70).

[487] Elias (n 303) 185.

identifying the amount of existing uncertainty about the data sharing practices of major data custodians, before reflecting on the impact of such fragmentation on independent researcher access. Finally, I will be concluding with the implications of this analysis for studying the formation of 'research data sharing institutions' in the UK public sector.

i. After the Administrative Data Research Network: fragmentation and the limits of formalisation

After its first years of operation, the Administrative Data Research Network (ADRN) was particularly struggling in facilitating access to ministerial department UK-wide data in such areas as social security, licensing and tax records.[488] In 2016, its Mid-Term review report noted that despite spending £25.3M (66% of its overall budget) it was 'behind its intended schedule for linked data projects, and far behind schedule for projects that made use of linked UK-wide data'.[489] ADRN's funding finished in 2018, as originally foreseen by the Economic and Social Research Council (ESRC), and its successor body, Administrative Data Research (ADR) UK, received ESRC funding on the condition that it would adopt a more effective approach than the ADRN in securing access to government data. During the last phases of its funding, the ADRN sought to demonstrate how its sponsored research can be more useful for government departments through so-called *thematic partnerships,* i.e. long-term relationships between the Network and specific data custodians in the interest of prioritising research on datasets and policy questions of central interest to government policy-makers.[490] The fundamental redirection of the Network's acquisition policy did pay off during the last year of its operation. Relationships between key government departments and the Network improved and data finally started becoming available for use by more ADRN-approved projects aiming to access government data.[491]

---

[488] Elias (n 303) 200.

[489] Research Project Management (n 29) 8.

[490] P Jackson, 'The Route to Better Knowledge for a Better Society' (2017) <https://prezi.com/6pe-8m2u9ad0/the-route-to-better-knowledge-for-a-better-society/?utm_campaign=share&utm_medium=copy> - these cover: children, world of work, ageing, productive society, and health and society.

[491] Strategic Data Negotiator, 'April 2018 Data Acquisition Project Report' (not published, on file with the author as part of his affiliation with the Administrative Data Service) 3. Note that the author's affiliation with Administrative Data Service did not play any role in the recruitment of participants, nor did it shape the research

The ADRN story is, arguably, a lesson in humility for 'pressurising constituents'[492] that seek to overcome established power imbalances through robust strategies of conflict. 'Efficiency', in the sense of satisfying data custodians that independent research can be of benefit to them as organisations, eventually mattered more than 'accountability' amidst increasing uncertainty about the requirements and procedures followed by data custodians in disclosing administrative data. Nonetheless, the ADRN, while it existed, played a significant role in extending research data sharing standards in the outer limits of the UK public sector. As a result, it is true that many transparent and formalised data sharing frameworks have been established, making 'the journey to obtaining data clearer and smoother' for independent researchers.[493] Indeed, an interviewee, who has worked for the Office for National Statistics, observes with satisfaction that:

> *'For a lot of the UK now, the idea that we should be default open is taken for granted, in a lots of places. Certainly, compared to what we started with many years ago, it's an awful lot easier to get access to data now.'*[494]

Indeed, in some cases, the ADRN has contributed to some data custodians being more 'friendly' to independent researchers, adopting open-ended definitions of the public good, allowing methodological studies without a clearly demonstrable policy impact[495] and responding to requests within strict timescales.[496] Such custodians, e.g. the Office for National Statistics in England or the Honest Broker Service in Northern Ireland, may also be more explicit about the availability of their datasets,[497] as well as less stringent in respect of their disclosure and confidentiality controls to maximise the usefulness of data for research.[498]

---

design in any manner so as to introduce risks related to bias. More broadly for this study's approach regarding mitigating bias see *supra*, chapter 3, D.

[492] Chowdury and Wessel (n 84) 340.

[493] Iveson and Deary (n 43) 97.

[494] Interview with participant nr. 42 (2020, virtual).

[495] Ritchie (n 71) 31.

[496] Jones and Elias (n 338) 51.

[497] Health and Social Care Northern Ireland (n 248).

[498] Ritchie (n 71).

This is not to say that existent data sharing frameworks are always relaxed in granting access when it comes to assessing the appropriateness of the applicant researchers or the contribution of a proposed research to the public interest. Some data custodians may be more likely to err on the side of caution, interpreting their published requirements narrowly and applying strict controls and safeguards. Custodians like the HM Revenue and Customs (HMRC) or the Department for Education, for example, restrict access or 'treat more sympathetically'[499] to requests for research that 'serves one of their functions'.[500] An interviewee who has worked for the HMRC concedes that this essentially confines the HMRC DataLab to things that are 'more akin to what the HMRC themselves would do'.[501] The Scottish government stipulates that it is 'crucial' for the facilitation of data access that a senior civil service manager or a minister endorses a project for its potential contribution to a 'national outcome'.[502] This means that certain bargaining and negotiation skills are still critical to securing access. Interested researchers, in these cases, should demonstrate how their outputs will benefit the custodian's understanding in specific policy areas e.g. modelling the effects of tax reforms or optimising economic efficiency in designing tax operations.[503] Even if that is the case, many frameworks are now formalised and, to an extent, transparent, allowing researchers to manage their expectations in respect of data access.

Are these 'many' frameworks enough to speak about a widespread formalisation of research data sharing in the UK public sector? While significant progress has been made, this is hardly the case. For a vast amount of data custodians, a fragmented picture emerges when seeking to identify the data sharing frameworks they are actively operating, with both major ministerial departments and 'smaller' public bodies or local authorities being silent about their preferable standards and procedures. A researcher working in Scotland concedes that major data custodians like the National Records of Scotland do not have 'any discoverability for their datasets'.[504] Another interviewee, who has worked for the Department for Education (DfE) in England and is now working for the National Foundation for Education Research (an

[499] Jones and Elias (n 338) 8.

[500] Bega (n 424) 62.

[501] Interview with participant nr. 32 (2020, Oxford).

[502] Scottish Government (n 406).

[503] Welpton and Wright (n 398) 7.

[504] Interview with participant nr. 29 (2020, phone).

independent research organisation), concedes that while they are familiar with the DfE's data sharing procedures, the same can hardly be said about other departments:

> *'Do I know how to do this with the Department for Work and Pensions? No (…) I don't know how long other government departments take (…) We have applied for data from Wales and Scotland while I'm here – that was hard work, very hard work, especially Scotland.'*[505]

Another interviewee with a similar background, i.e. first a DfE analyst and now working for the Education Policy Institute (an independent research organisation), complains about lack of clarity in the DfE's own procedures. Especially in the beginning of their interactions with the DfE as external researchers, they found the DfE guidance as 'quite limited' and 'information does not really flow appropriately many times'.[506] Interviewees with experience in negotiating data access across the UK corroborate this impression of uncertainty. A researcher working in Scotland finds the approval procedures in the Scottish public sector as 'extremely messy' and 'very bumpy',[507] whereas an interviewee working for an intermediary body in Scotland claims that 'disparity among data providers and how they do things' is even greater when trying to access central government data.[508] Another interviewee working as an administrative data researcher in Scotland associates lack of clarity with lack of experience in developing data sharing policies. In their view, transparency around data sharing is 'hit and miss' with smaller organisations.[509] An interviewee who has worked for the Office for National Statistics makes a counter-argument to that, stressing that officials in large data custodians are 'strongly incentivised' to refrain from establishing and enabling transparent data sharing procedures.[510]

Other interviewees caution against the meaningfulness of this distinction, highlighting that, with larger organisations, there is a lot of internal disparity in how data sharing works. Departments like the Department for Education are 'complex organisations, with lots of

[505] Interview with participant nr. 30 (2020, virtual).

[506] Interview with participant nr. 33 (2020, virtual).

[507] Interview with participant nr. 34 (2020, virtual).

[508] Interview with participant nr. 36 (2020, virtual).

[509] Interview with participant nr. 37 (2020, virtual).

[510] interview with participant nr. 42 (2020, virtual).

different parts'.[511] An interviewee working for the Welsh government makes this point when referring to their own department:

> *'The thing about Welsh Government is that even Welsh Government is not a homogenous whole (…) parts that came in from the wider public sector had a different perspective, a different history. (…) surely, there are still pockets of Welsh Government that would not know data linking if you kicked them on the head.'*[512]

Nonetheless, the lack of procedural clarity becomes more striking when thinking about the major ministerial departments with significant organisational capacities and a wealth of administrative datasets at their disposal. For example, HM Courts and Tribunals, while in the midst of an ambitious digitalisation reform that will generate a wealth of novel justice-related datasets, delegate their research data sharing decision-making to an informal data access panel which communicates via e-mail and does not record or publicise minutes of meetings or particular access requirements.[513] The Department for Work and Pensions (DWP), one of the pioneer bodies in government-led administrative data research,[514] are also laconic about their research data sharing procedure, encouraging researchers to submit a written application to the department and avoiding to publish information about the progress of applications online.[515] Transparent and formalised data sharing frameworks, similarly, shine through their absence on the websites of most local authorities or other public bodies.

As described by an interviewee with work experience in the Administrative Data Research Centre in Northern Ireland, the Administrative Data Research Network's efforts to formalise research data sharing was intended as a '*push factor*' but was perceived by many departments

[511] interview with participant nr. 40 (2020, virtual).

[512] interview with participant nr. 25 (2019, Cardiff).

[513] Byrom (n 428) 6; HMCTS (n 190).

[514] Department for Work and Pensions, 'DWP Open Data: Case studies' (2013) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/221159/dwp-open-data-case-studies.pdf>; National Archives, 'Information Management Assessment: Department for Work and Pensions' (2015) <https://www.nationalarchives.gov.uk/documents/information-management/dwp-ima-report-2015.pdf>.

[515] Department for Work and Pensions (n 337).

as a '*fear factor'*.[516] Beyond merely neglecting to disclose information about their data sharing practices, some data custodians actively resist providing any relevant information to the general public. The DWP recently rejected a Freedom of Information Act (FOIA) 2000 request to disclose information about their use of their customers' data.[517] The HM Revenue and Customs responded to a similar request by providing only partial information about their external data sharing for research purposes and refusing to disclose information about their internal procedures. As a result, in spite of the significant progress in formalising independent research access to government data, it currently seems 'unlikely that a standard pattern of use for researchers in general would emerge across government'.[518] In many cases, data access seemingly remains 'unregulated, unreliable and haphazard',[519] with researchers facing the tall task of navigating a complex landscape without a reliable compass in their possession. This seems to sketch a picture that is often present in power struggles: both conflict and bargaining is helpful to ameliorate power imbalances, yet this is often not enough to radically eradicate such imbalances, at least for a long time. Who are, then, the 'authors' of the applicable data sharing standards and practices within this uncertain and fragmented picture?

ii. The eclipse of 'accountability' and the complications for independent researchers

While elements of a popular narrative throughout the development of this space, 'accountability' and transparency of data sharing procedures seemingly waver in practice. In the absence of comprehensive knowledge about the data sharing standards and procedures that are applied by most data custodians in the UK public sector,[520] commentators and stakeholders have drawn on anecdotal information or individual researcher narratives to reflect on the

[516] Interview with participant nr. 10 (2018, Belfast).

[517] Booth (n 62).

[518] Smith et al (n 328) 28.

[519] Elias (n 303) 184.

[520] Smith et al (n 328) 39: 'there are no consistent patterns or always clear reasons for the different levels of access that different groups have obtained'; Iveson and Deary (n 43): 'without single points of administrative data permissions, researchers must negotiate with and complete the governance procedures of all concerned data controllers, resulting in a complex and often unclear process'.

governing forces of data access.[521] According to Elias, most custodians prioritise such factors as the legality of the request, the public interest, the resources they will need to allocate in meeting the request and the risks they would be exposed to if they grant access.[522]

As resources and disclosure risks become increasingly pertinent, the ideal of bolstering the 'accountability' of government towards its citizens through independent research on administrative datasets breaks down, with more efficiency-driven rationales becoming more prominent. With data custodians being likely to approve requests which do not generate risks of portraying them negatively or critically,[523] Administrative Data Research (ADR) UK, the Administrative Data Research Network's successor, has opted to facilitate research that can 'predominantly help inform (…) government policy decisions (….) ensuring that decision making is evidence based'.[524] Researchers are called to 'compromise on the detail of the data' and potentially 'bundle together a number of projects' in the interest of satisfying the appetite of government departments.[525] Access to administrative data is often being regarded as a 'privilege, not a right'.[526]

Various interviewees shed light on the prevalence of 'efficiency' within custodians' data sharing policies. An interviewee working for the Department for Work and Pensions notes that their internal review board, called the External Data Sharing and Access Committee, assesses the 'usefulness (…) (ie) the practical use' of the proposed research within their definition of the 'public benefit'.[527] An interviewee who has worked for the HM Revenue and Customs (HRMC) also notes that when defining their 'prioritisation' criteria within the HMRC data sharing policy main considerations would relate to 'higher priority and better fit with HMRC objectives', as well as 'ease to deliver'.[528] Another interviewee working for the Office for

[521] E.g. Iveson and Deary *ibid.*

[522] Elias (n 303) 194.

[523] *Ibid* 196.

[524] O'Reilly et al (n 351).

[525] Jackson (n 308) 24.

[526] J Lane, 'Building an Infrastructure to Support the Use of Government Administrative Data for Program Performance and Social Science Research' (2017) 1 The ANNALS of the American Academy of Political and Social Science 240, 246.

[527] Interview with participant nr. 8 (2017, Sheffield).

[528] Interview with participant nr. 32 (2020, Oxford).

National Statistics attributes this imbalance between 'efficiency' and 'accountability' to the fact that many policy-making departments feel that 'the more they would allow the public to see, the harder it is going to be for them to actually put policies in place'.[529] Where does this leave independent administrative data researchers?

Power imbalances between researchers and data custodians are as clear as ever. Much like the first phase of administrative data sharing for social-science research, researchers have to grapple with a 'complex and changeable' landscape within which they are unable to predict how long they will need to access data and how they should fulfil data provider requirements.[530] Such requirements are often communicated in a way that is not helpful to researchers or relate to technical and legal information governance issues that they are not necessarily able to comprehend.[531] A researcher working in Scotland recalls their surprise when, after an initial approval from the data custodians, they were asked to undergo a 'whole new process' of 'negotiating the contracts (and) (…) the particular conditions of access etc'.[532] Another interviewee who has worked for the Department for Education concedes that the expansion of data sharing policies to novel areas may take long and make the lives of researchers difficult:

> *'If you do it for the first time, then good luck, I don't think you will necessarily navigate this, at least not after a long time.* '[533]

The task is even taller for researchers who are interested in linking datasets managed by more than one data custodian. Without sufficient clarity on the compatibility between the respective data sharing frameworks, researchers are often expected to conform with conflicting rules and guidance on meeting various data sharing requirements.[534] An interviewee who has been attempting to link datasets between different parts of the Scottish government reflects on their so-far unsuccessful efforts over the last six (6) years:

---

[529] Interview with participant nr. 3 (2017, Titchfield).

[530] Iveson and Deary (n 43).

[531] *Ibid.*

[532] Interview with participant nr. 29 (2020, phone).

[533] Interview with participant nr. 33 (2020, virtual).

[534] G Laurie and N Sethi, 'Information Governance Of Use Of Health-Related Data In Medical Research In Scotland: Current Practices And Future Scenarios' (SHIP Core Programme 2 Working Paper No. 1) 9.

> *'To be honest, it's been such a long time, I'm not entirely sure what the roadblock was other than the fact that we had to have the right sign-off from (…) as part of their internal sub-linkage things.'*[535]

Under these circumstances, access becomes a time-demanding and herculean task, whereas disclosure decisions are hardly consistent across government departments in respect of the application of access rules. As a result, fewer researchers are inclined to embark on administrative data research, observing that many of their colleagues had to abandon existing plans and funding due to the data not arriving in time (or at all). An interviewee working in Scotland goes as far as conceding that they actually amended their research agenda in the light of data sharing considerations:

> *'Let's say that it was not the original focus we were interested in, it was part of a wider research program (…) In the course of the applications process, we had to (…) change the focus.'*[536]

The secondary literature is also awash with researcher narratives of long delays and struggles in getting access to administrative data. Iveson and Deary report long delays and uncertainty in trying to use data from the Scottish Census and link it to health data.[537] While they found health data permissions and access procedures in Scotland much more streamlined, they faced various complications such as a change in the legal interpretation of the power to share data by the Scottish government or an unpredictably long queue to have their request considered.[538] Eventually, they had to wait more than three years to acquire access to linked data, also opting for a more modest choice of datasets than originally envisaged.[539] Morris and colleagues offer a lucid account of similar difficulties while trying to link datasets belonging to the Department of Education (the National Pupil Database) and the National Health Service Digital (Hospital Episode Statistics).[540] The research team negotiated access with data custodians over three (3)

[535] Interview with participant nr. 35 (2020, virtual).

[536] Interview with participant nr. 38 (2020, virtual).

[537] Iveson and Deary (n 43).

[538] *Ibid.*

[539] *Ibid.*

[540] H Morris et al (n 114) 97; the time mentioned here does not include the additional time within which technical challenges of data linkage were dealt with.

years and eleven (11) months, conducting six (6) meetings and one-hundred and eight (108) email and phone correspondences with the custodians before an agreement on linkage of datasets could be reached.[541]

Other researchers corroborate that they faced a 'complex landscape, unclear and conflicting guidelines',[542] lamenting the 'extraordinarily disturbing' impact on the careers of academics, particularly early career researchers who will produce a reduced research output as a result.[543] These narratives are telling of power imbalances between researchers and data custodians. They are indicative of the strategic changes the Administrative Data Research Network had to implement to increase the likelihood of satisfying data custodian requirements and achieving greater progress in data acquisition. The last sub-section takes stock of the central lessons learned from the historical development of this space for the purposes of empirically studying the formation of research data sharing institutions in the following chapters.

D. Towards studying research data sharing institutions: institutions & interactions

This chapter has analysed how the continuous employment of both conflict and bargaining strategic approaches between researchers and data custodians has been recalibrating the UK research data sharing space in the last four decades. From the dawn of 1980s to the dusk of the 2010s, the governance of administrative data sharing for research purposes in the UK public sector seems to have come full circle. It started as an ad-hoc, informal activity, initially being led and conducted within the confines of public authorities and gradually expanding to a closed circle of entrusted academics until the 2000s. With various actors in the data sharing space evangelising about the importance of widening access to independent social-science researchers for 'efficiency' and 'accountability' purposes, formal data sharing frameworks with concrete requirements and approval stages were established and their use became more widespread due to initiatives like the ADRN. Yet, as we are moving forward, the process of data sharing remains fraught with significant uncertainty, with major public-sector data

[541] *Ibid.*

[542] Brett and Deary (n 474).

[543] Iveson and Deary (n 43): 'the current situation essentially prohibits researchers, particularly early-stage researchers from conducting projects involving linked administrative data unless permissions are sought well in advance'; Sexton et al (n 380).

custodians avoiding or actively resisting to publicise their criteria and requirements for granting access. With this analysis of the actors of the space in mind, a few central questions arise for the formation of the 'rules of the game' that filter access to government data for social-science research purposes. These questions emanate from the overarching view of the space provided in this chapter and seek to address more specific issues that relate to the other two dimensions of the presently adopted analytical framework, i.e. institutions and interactions.

First, to what extent do such research data sharing institutions as the complex technicalities and legalities of data sharing determine the creation and application of more or less transparent and predictable frameworks in practice? Going beyond the processes of bargaining and conflict between the actors presented in this chapter, this question hints at the substance of the 'rules of the game' filtering access to administrative data for research purposes in the UK. As so-far analysis has suggested, data sharing for research is a technically complex and multi-layered activity, within which data custodians and requesting parties need to reconcile various considerations. For example, they need to balance the public interest (e.g. in 'economic growth' or 'scrutiny of public power') with such other interests as their organisational priorities, or the privacy and data protection rights of data subjects. The UK research data sharing literature is sharply divided, however, on the extent to which legal requirements of e.g. technical security and data minimisation actually govern research data sharing,[544] as opposed to such other factors as organisational culture.[545] This divide is also reflected in the data presented in this chapter, with such actors as researcher groups at times becoming adversarial and invoking legal principles to justify their rights to access more government data, whereas, at other times, prioritising a more strategic bargaining approach and seeking to build rapport and trustworthy relationships with data custodians.

Second, how do the interactions between the actors of this space, mainly government departments, researchers and intermediary bodies, shape the actually applied requirements and processes for data access? This question goes beyond the historical development perspective presented in this chapter and seeks to interrogate the interactions shaping the nuts-and-bolts of research data sharing institutions e.g. the establishment of particular organisational criteria of

[544] Elias (n 303) 191: 'This was not a consequence of the different legal systems in these countries, but more a reflection of their smaller and more integrated administrative departments of government'.

[545] E.g. Thomas and Walport (n 2); Laurie and Stevens (n 7).

data access or the holding of certain collective beliefs that may structure individual data-sharing decision-making. Some initial directions to respond about that have been offered in this chapter. The historical account presented here has indicated that changes in this space are associated with, if not determined by, the perpetuation of specific narratives and the gradual development of confidence in the merits of research data sharing. Relationships and communications between different actors are integral to influencing developments, leveraging resources and asserting authority. This, at times, results in a widening, while at other times in a narrowing of access to data for independent social-science research. It will be for the following three chapters to elaborate on these questions and articulate empirically-informed and context-specific arguments about the 'rules of the game' that govern UK research data sharing.

5. THE LIMITED REACH OF 'LEGAL' INTERPRETATIONS OF ACTORS IN THE REGULATION OF RESEARCH DATA SHARING DECISIONS

> *'The premise we have reached so far is that legally we can link it (…) but socially do you want to? How many times would someone (…) find that this data has been linked behind their backs without their consent? Socially, how long you can do that for?'* [Interviewee 10]

> *'What tends to happen is you make the decision, you then look at the Regulations and see whether you can back that up with what the Regulation says.'* [Interviewee 23]

> *'So, to argue that this is a solid legal foundation is pretty much nonsense (…). But, what it was, it was definitely within the spirit of working for the public good. (…) The public benefit, the spirit of the public benefit, trumps the spirit of the law. You deal with the letter of the law and the spirit of the public benefit.'* [Interviewee 42]

The previous chapter tracked the historical development of research data sharing regulation in the UK public sector, sketching the ways in which organisational actors sought to influence the construction and reconstruction of the space through the perpetuation of certain narratives. As it was noted,[546] the different constituents of this space have invoked a plethora of factors, structured around the narratives of 'efficiency' and 'accountability'. These factors range from policy-making considerations and resource-related limitations to the so-called 'technicalities and legalities' of data sharing. Which factors, however, are the most salient ones in the actual construction and application of data sharing frameworks in practice? Which are the 'rules of the game' that filter access to administrative data for research purposes on the ground?

This chapter shifts the analysis from actors to the formation of research data sharing institutions, particularly interrogating the role of 'law' and legal interpretations as a driving force in data sharing regulation. It argues that interpretations of legal powers to share data for research purposes and legal limitations on such powers are relatively weak in influencing the

[546] *Supra* chapter 4, A-C.

establishment or application of data sharing frameworks in practice. As it will be shown in subsequent chapters, the most salient factors shaping research data sharing institutions stem from a collective 'research data sharing culture'. They consist in unwritten and uncodified conventions, assumptions and normative attitudes towards the ownership of administrative data and the relevance of research for policy-making.

Law's limited reach in influencing the regulation of research data sharing in the UK public sector is illuminated by a presentation of interview data from the four nations: England, Northern Ireland, Wales and Scotland. Research data sharing participants provide accounts that are not uniform, yet converge on the understanding that 'legal' considerations are not predominantly shaping organisational responses to research data sharing regulation. Before presenting these empirical findings, the chapter elaborates on how 'legal' interpretations are to be construed within this study. The 'legal' is defined proceeding from the formal legal framework as the starting point and extending it so as to encompass my participants' perceptions of legal powers and limits on research data sharing.[547]

A. 'Legal' Interpretations as Research Data Sharing Institutions

There is a need to briefly clarify from the outset both how 'legal' interpretations are used analytically in this study and how they form part of 'research data sharing institutions' as they have been defined in chapter 2.[548]

With regard to the first issue, 'law' and the 'legal' have been employed within socio-legal studies in multiple ways, ranging from 'legality' as an ongoing social activity to positive/State law.[549] Legal consciousness studies are well-known for developing broad definitions of the 'legal', conceiving legality as a social structure constituted *in* society, rather than acting upon society as an external entity. Using the notion of *legality* in general, such a definition

[547] This is coined as an 'interpretive' definition of the 'legal' in socio-legal studies, see D Cowan, 'Legal Consciousness: Some Observations' (2004) 67 Modern Law Review 928, 942.

[548] *Supra* chapter 2, C.

[549] P Ewick and S Silbey, 'Narrating social structure: stories of resistance to legal authority' (2003) 108(6) American Journal of Sociology 1334; *cf* D Cooper, 'Local government legal consciousness in the shadow of juridification' (1995) 22(4) Journal of Law and Society 506-526; S Halliday, *Judicial Review and Compliance with Administrative Law* (Hart 2004).

accommodates subjects' perceptions of both state and non-State law as relevant to its construction. Even a mundane everyday life activity (e.g. 'placing chairs in shovelled out parking spots') is perceived as contributing to the social construction of legality *if* it is interpretable by an audience in exclusive connection with legal concepts and terms (e.g. property, labour etc).[550] At the other end of the definitional spectrum, some socio-legal scholars have proceeded from a much narrower notion than legality in general: *positive State law* or specific aspects of positive State law. These scholars have been traditionally interested in so-called 'implementation' or 'gap' studies about the empirical realities of 'law in action' as opposed to 'law in the books'.[551]

Between these two approaches, scholars have been developing definitions of the 'legal' which proceed from the formal legal framework but extend to law as research participants 'receive and perceive it'.[552] Within these definitions, official, State law usually provides a crucial reference point. The aim of such definitions is to avoid the weaknesses of both overly inclusive definitions, which perceive law's social empire as all-encompassing,[553] and overly restrictive definitions, which subscribe to the much-criticised analytical privileging of official law.[554]

The present thesis adopts such an *interpretive* definition of the 'legal', in the interest of interrogating the relevance of 'legal' *interpretations* of data sharing powers and limitations for the regulation of research data sharing decisions. In the first place, 'legal' interpretations are the ones related to legal powers and limits, as prescribed by State law. Through empirical work, however, this definition extends to cover this study's participants' perceptions of what they *understand* that is legally required from them or what they *feel* are legally enabled to do. In other words, the formal legal framework operates as a filter to ensure that the notion of legality adopted by a study is not too broad and all-encompassing. At the same time, it is not the existence of e.g. a power in official law that makes it part of 'legality', but the *perception* and

[550] S Silbey, 'Studying Legal Consciousness: Building Institutional Theory from Micro Data' (2018) 3 Droit et société 698, 725.

[551] S Halliday, 'After Hegemony? The Varieties of Legal Consciousness Research' (2019) 28(6) Social & Legal Studies 859.

[552] Cowan (n 540) 175.

[553] See the critique in M Hertogh (n 162).

[554] Silbey (n 543); S E Merry, 'Legal Pluralism' (1988) 22 Law and Society Review 869.

interpretation of this power as 'legal' by the study's participants. In that sense, an interpretive definition is different both from the broader (which disregard formal law as an important definitional consideration) and the narrower (which confine themselves to formal law) definitions of the 'legal'.

Adopting this definitional viewpoint is meaningful not only by reference to this study's analytical and methodological commitments,[555] but also by virtue of the particular characteristics of the research data sharing space. Adopting an interpretive definition of the 'legal' does recognise that stable, settled and institutionalised legal norms are important, without negating the agency of subjects to renegotiate the contents of the law through their daily interactions. As chapter 4 has demonstrated, the research data sharing space has been constantly recalibrated based on what was regarded as possible and desirable by the actors populating it. The perceptions of these participants may differ from an expert analysis of data sharing law.[556] Yet, such institutionalised formal data sharing laws, e.g. the Digital Economy Act 2017, still operate as a common frame of reference for actors in this space, considering that research data sharing is a heavily routinised and technical activity that is subject to a plethora of legal and information governance requirements.

Having said that, how are we to demarcate the 'legalities' of data sharing from other relevant considerations which structure the process of decision-making? Chapter 2 defined research data sharing institutions as a broad church, including codified rules, organisational policies, norms, conventions, customs and cultures.[557] In research data sharing, legal considerations may be intertwined with other factors e.g. technical requirements. For example, issues related to data security are both legal, concerning the conformity of data participants with the data minimisation principle,[558] and technical, in the sense that certain systems architecture solutions

[555] *Supra* chapters 2, A-C and 3, where the constructivist-interpretative analytical premises underpinning the regulatory space analysis and qualitative methodological lenses adopted in this study were articulated.

[556] As the following discussion will demonstrate, it is often the case that the nuances or the exaggerations that are reflected in the narratives provided by this work's interviewees do not always match the findings of a doctrinal legal analysis, see e.g. *infra* p 146.

[557] *Supra* chapter 2, C.

[558] See art. 5(1)(c) GDPR on the obligation of data processing to be 'adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (data minimisation)'.

need to be implemented in building a secure data sharing procedure.[559] Information governance officials,[560] in fact, may often be tasked with resolving both kinds of questions within a data sharing collaboration.

Yet, the academic and policy literature on research data sharing in the UK has been sharply divided on whether 'legal' considerations have been more salient than other factors, e.g. organisational culture, in shaping organisational policies regarding data access by researchers.[561] This chapter interrogates literature assumptions by assessing the role of 'legal' interpretations in shaping research data sharing decisions. It does so by proceeding from an overview of the formal legal framework governing this activity,[562] before presenting empirical data on the role of law within research data sharing decisions. It is true, as chapter 4 demonstrated,[563] that actors in the research data sharing space have been invoking 'legal' reasons, e.g. uncertainty about the legal power to share data, to justify the implementation or lack thereof of particular data sharing frameworks.[564] As empirical data presented in this chapter will demonstrate, however, the extent to which these considerations actually determine the construction of research data sharing institutions is rather limited in practice.

B. The Legal Framework applicable to Research Data Sharing in the UK

Legal regulation can either be *restrictive*, seeking to limit the actions of regulated entities and ensure their compliance with certain rules, or *facilitative*, seeking to empower regulated entities to engage in certain types of activity, or both.[565] In the case of research data sharing, legal regulation is in principle facilitative, since the particular type of data sharing is always

[559] J Campos et al, 'The Challenges of Cybersecurity Frameworks to Protect Data Required for the Development of Advanced Maintenance' (2016) 47 Procedia CIRP 222. See e.g. Jones and Elias (n 338) 39 on the Office for National Statistics Virtual Microdata Laboratory.

[560] *Supra* chapter 3, A on the role of information governance officials.

[561] Law Commission (n 2); Administrative Data Taskforce (n 23); *cf.* Thomas and Walport (n 2); Laurie and Stevens (n 7).

[562] This includes both an analysis of formal legal powers to share data, as well as restrictions on such powers in UK and EU data protection and human rights law.

[563] *Supra* chapter 4, A-C.

[564] *Ibid.*

[565] Morgan and Yeung (n 54) 75.

discretionary and not mandatory,[566] combined with a number of restrictions to ensure compliance with privacy and data protection safeguards. Data custodians are enabled, by law, to engage in data sharing for a number of legitimate purposes, including social-scientific research. To provide a starting point for the assessment of 'legal' interpretations as a factor influencing the construction of research data sharing institutions,[567] this section provides an overview of the formal legal framework applying to this activity in the UK. This overview is distinguished into two parts: legal powers to share data for research purposes and limitations on such powers.

A few caveats need to be registered here before embarking on the analysis. First, this is an area of law that is largely under-theorised and untested in judicial review. While the law of administrative data sharing *simpliciter* has been discussed time and again in the literature,[568] the same cannot be said about the sharing of data *for research*.[569] As a result, some of the findings of this overview do not benefit from being tested in judicial practice. Second, this overview does not exhaustively analyse specific provisions that only apply to one public body, e.g. the Prescribed Persons Act 2009 permitting the use of personal data held by the Department for Education for research purposes. Although such specific instruments will be mentioned, emphasis will be laid on the general, as in applicable to all public bodies, legal instruments covering data sharing like the Digital Economy Act 2017 and the Data Protection Act 2018. Third, data sharing legislation is, by and large, applicable to all four UK countries.[570] Being consistent with the second caveat, the analysis will be confined to UK-wide legislation which specifically covers data sharing, with local legislation only mentioned when it raises an issue of general importance.[571]

[566] Department for Culture, Media & Sport (n 28).

[567] *Supra* chapter 2 on 'research data sharing institutions' as the 'rules of the game' that filter access to administrative data for research purposes in practice.

[568] Law Commission (n 2) [49-100], Thomas and Walport (n 2) [13-26]; Laurie and Stevens (n 7) 11.

[569] There are, of course, exceptions, such as Stevens's work on the implications of the EU GDPR for social research; Stevens (n 120) 97.

[570] The Data Protection Act 2018 and the Digital Economy Act 2017.

[571] E.g. the Adults With Incapacity (Scotland) Act 2000 (AWI(S) Act 2000) or The Education (Pupil Information) (Wales) Regulations 2011 specify the application of the overarching legislation in local contexts, but do not introduce distinct legal principles.

i. Powers to Share Data for Research Purposes in UK Law

Legal powers to share data stem from a variety of 'sources': statutory gateways, implied, common-law and prerogative powers.[572] The analysis presented here suggests that not all of these sources are equally applicable in the research-data-sharing context; apart from statutory gateways, common-law powers might be a more appropriate source than implied powers, whereas prerogative powers are highly unlikely to be applicable.

a. *Statutory gateways*

The most uncontroversial legal source of data sharing powers is (administrative) statutory law. In general, there are many different and context-specific legal powers to share administrative data deriving from statutory acts.[573] Actors in this space, e.g. the Department for Work and Pensions, are subject to different pieces of legislation and have actually used such legislation to engage in research data sharing activities, particularly at the early stages of this space's historical development.[574] To give a few examples, key 'gateways' are to be found in sections 251 of the National Health Service (NHS) Act 2006, 39 of the Statistics and Registration Services Act (SRSA) 2007, 261 of the Health and Social Care (HSC) Act 2012, 3 of the Prescribed Persons Act 2009 and 64 of the Digital Economy Act (DEA) 2017.

Until the recent enactment of the DEA, these gateways empowered only a few public bodies to share data for research. For example, the 'approved researcher' gateway in the SRSA empowers the Office for National Statistics to share data for statistical research purposes. The NHS and HSC Acts are similarly limited in the scope of research allowed. Section 251(12) of the former confines its use to *medical* research; section 261 of the latter, after its amendment by section 122 of the Care Act 2014, requires that the purposes of data sharing are limited to the 'promotion of health', i.e. proposing improvements in the operation of the health and social care system.

[572] Law Commission (n 2) part 2.

[573] E.g. sections 47 of the Children Act 1989, 175 of the Education Act 2002, 2 of the Local Government Act 2000 and 17 of the Crime and Disorder Act 1998.

[574] *Supra* chapter 4, A-B.

The enactment of the DEA in April 2017 was a step-change to this piecemeal approach. In fact, the Administrative Data Research Network investment, which was described previously as a torchbearer in the formalisation of research data sharing in the UK,[575] hailed the enactment of this law as a 'historic moment' for administrative data sharing for research in the UK.[576] Section 64(1) of the Act empowers all public authorities to disclose data they hold 'in the context of their functions' for research purposes. Public bodies may disclose such data in accordance with the conditions outlined in sections 64(3)-(10),[577] as well as with the Data Protection Act 2018 which implements the EU General Data Protection Regulation into domestic law. While section 73(2) of the Act excludes information held relating to health and social care functions, the discretionary power established by section 64(1) is now the major source for most research-data-sharing purposes. What happens, however, in cases not covered by explicit statutory language, e.g. when public authorities are unsure about the nexus between the requested datasets and their functions?

b. *Implied powers*

An answer to the above question may be given by the so-called 'implied powers' of public bodies. The Law Commission maintains that in the absence of an explicit legal gateway, the power may be implied if 'data sharing is *reasonably incidental* to an express power to do something else'.[578] The problem is that there has been limited discussion on implied powers in the research context. In fact, one struggles to find a single legal resource or precedent on implied powers as the legal basis for research data sharing. The 2014 Law Commission report, for example, mentions several examples of implied powers, emerging e.g. from a duty to co-operate or ensure the proper functioning of a regulatory body; none of them are related to research.[579] This might explain the reluctance of many data custodians in the research data sharing space, such as the Northern Irish Statistics and Research Agency or the Department for Work and Pensions, to rely on implied powers.[580]

---

[575] *Supra* chapter 4, B, iii.

[576] ADRN (n 13).

[577] Mainly the removal of identifying information and adherence to the requirements to be published by the UKSA via its 'Code of Practice' pursuant to section 70 DEA 2017.

[578] Law Commission (n 2) [1.35].

[579] *Ibid* 1.41 *et seq.*

[580] Interview with participants nr. 7, 8 (2017, Sheffield) and 9 (2018, Belfast).

Similarly unenlightening for research data sharing is the picture in examples of guidance produced by major stakeholders. The Centre of Excellence for Information Sharing suggests examining whether an activity is 'reasonably associated' with the existing powers of a public body in identifying an implied power to share data.[581] The Information Commissioner's Office advises a consideration of: the type of organisation, its relevant functions or powers, the nature of the requested information (e.g. confidential or not) and any legal obligation to share it.[582] While the above are clearly helpful on a general level, there is no suggestion that 'research' is an activity that could be considered relevant to the 'function' or 'type' of specific public bodies.

This lack of precedents and guidance invites an enquiry into the scope of the 'reasonably incidental' norm. Lord Selborne LC defined it in *Great Eastern Railway* as 'whatever may fairly be regarded as incidental to, or consequential upon, those things which the Legislature has authorised'.[583] This interpretation of implied powers has been time and again endorsed by the courts,[584] even in the data sharing context. In *R(W),* the Administrative Court was asked whether National Health Service (NHS) bodies have the legal power to disclose non-medical information of non-resident debtors for health service charges to the Health Secretary. The latter would later inform the Home Secretary who would apply immigration sanctions. Even though an explicit legal provision was not in place, Mr Justice Silber found that such a legal power existed 'under the general powers of NHS foundation trusts to do anything (…) necessary or expedient (…) in connection with their functions' and the Secretary's 'general function of promoting an effective healthcare system'.[585]

Is the above dearth of legal resources on 'implied powers' and research data sharing, then, merely a coincidence? One is inclined to answer in the negative by looking at the 'reasonably incidental' interpretation by the courts. Specifically, it was ruled in two major Supreme Court

---

[581] Centre of Excellence for Information Sharing, 'Information Governance' <https://informationsharing.org.uk/infogovernance/>.

[582] ICO, 'Data Sharing Checklists' <https://ico.org.uk/media/for-organisations/documents/1067/data_sharing_checklists.pdf>.

[583] *Attorney General v Great Eastern Railway Co* (1880) 5 App Cas 473.

[584] *D and J Nicol v Dundee Harbour Trustees* (1914) 2 SLT 418*; Glasgow Corporation v Flint* (1966) SLT 183.

[585] *R (on the application of W, X, Y, and Z) v The Secretary of State for Health v The Secretary of State for the Home Department* (2014) EWHC 1532 (Admin) (Silber J) [82].

judgments, i.e. in *Ward* and *New London College*,[586] that only considerations which contribute to or emanate from the expressly stated main function of a public body may give rise to an implied power. Their Lordships did not agree on the matter; more restrictive approaches such as the one followed by the majority in *Ward* require that the 'implication has to be *necessary* to make the statutory power effective to achieve its purpose',[587] whereas the minority in *New London College* demanded that 'the source of the incidental power was found in a *specific provision* conferring specific functions'.[588] Even relaxed approaches, however, require a connection between the incidental action and the body's main function.[589]

Demonstrating that research-data-sharing is at least related to, if not necessary for, a public body's core function might prove challenging for most data owners.[590] This is the case, at least, within models like the Administrative Data Research Network, where research data sharing was carried out on an 'individual researcher' basis and was not conditioned upon producing outcomes that will help departments improve their policy-making.[591] Even if research data sharing is being undertaken to serve the 'public interest', the onus of convincingly connecting this notion with the specific function of each owner-public body may urge data custodians to look for other legal options.

c. *Common law powers*

The 'administrative' common-law powers of the Crown present such a solution. These powers are limited to data custodians who are ministerial departments. These custodians do not only hold the most comprehensive administrative datasets in the UK, but have also mostly pioneered

[586] *Ward v Commissioner of Police of the Metropolis and another* (2005) UKHL 32; *R (New London College Ltd) v Secretary of State for the Home Department* (2013) UKSC 51.

[587] *Ibid Ward* (Baroness Hale) [24-27].

[588] *New London College* (n 579) (Lord Carnwath) [33-36].

[589] *Ibid* (Lord Sumption) [29]: 'If the Secretary of State is entitled to prescribe (…) rules for the grant of leave (…) which demand upon the migrant having a suitable sponsor, then she must be also entitled to take administrative measures for identifying sponsors (…)'; *Ward* (n 579) (Lord Rodger) [5]: 'those things which are incident to, and may reasonably and properly be done under the main purpose (…) would not be prohibited'.

[590] Even if, for a few special cases like the Office for National Statistics, whose functions include promoting research, meeting this threshold may be straightforward.

[591] The Network is currently discussing a change of policy to tempting data custodians with a connection between research projects and their own policy work, see Jackson (n 483).

the introduction of research data sharing initiatives in the UK public sector. As it will be recalled from chapter 4,[592] such custodians as the Department for Education, the Department for Work and Pensions and HM Revenue and Customs have been among the first ones which commissioned academic researchers to perform analyses on their datasets.

The doctrinal legal analysis of common law powers is arguably more straightforward than implied powers. Lord Sumption noted in *New London College* that 'the Crown possesses some general administrative powers (…) which are not exercises of the royal prerogative and do not require statutory authority'.[593] Perry describes such powers as 'unglamorous, but important', accounting for 'much of the ordinary business of the government'.[594] This category encompasses such powers as the power to form contracts, conveying property, making *ex gratia* payments, appointing agents and creating policies.[595] While administrative datasets are not 'property' of the public bodies,[596] the latter are, by law, responsible for their use and management. Allowing access based on data sharing arrangements to serve a policy promoting the public interest arguably falls within this ambit. Wade and Forsyth offer another relevant argument; section 1 of the Localism Act 2011 provides local authorities with 'the power to do *anything* that individuals generally may do' (emphasis added).[597] An individual shares data for research lawfully, if compliance with specific legal restrictions is satisfied. In case local authorities have this power, then, it would be surprising for central Government to lack it.[598] This is a better argument than the contestable contention that the Crown, as a corporation sole, possesses all the legal powers that a natural person does.[599]

[592] *Supra* chapter 4, A-B.

[593] *New London College* (n 579) (Lord Sumption) [28].

[594] A Perry, 'The Crown's administrative powers' (2015) Law Quarterly Review 652.

[595] *ibid* 661 citing: *Rederiaktiebolaget Amphitrite v The King* (1921) 3 KB 500 (Rowlatt J) [503]; *Robertson v Minister of Pensions* (1949) 1 KB 227 (Denning J) [231]; *Crown Lands Commissioner v Page* (1960) 2 QB 274 (Lord Evershed MR) [287-288].

[596] *Supra* chapter 4, C, ii.

[597] Section 1 of the Localism Act 2011. The Act sought to devolve decision-making powers from the central government in England and Wales to individuals and communities.

[598] W Wade and C Forsyth, *Administrative Law* (OUP 2014) 180.

[599] Perry (n 587) 658 identifies key deficiencies in this argument.

There are, however, two caveats to this conclusion. First, many departments, with the Department for Work and Pensions (DWP) being a prime example, claim that their common-law powers have been 'eroded' by the proliferation of statutory legislation.[600] In this view, a ministerial department cannot use common-law powers to cover the gaps of statutory legislation, since the latter exclusively regulates specific areas of decision-making and 'eats away' the common-law. While the idea of common-law 'erosion' is indeed well-established,[601] its scope is far from clear. For instance, in *Hooper,* the House of Lords refrained from ruling on whether a statutory requirement to make payments to widows 'eroded' the common-law power of the DWP to make payments to widowers.[602] Therefore, public bodies may find that they possess such powers contingent on the circumstances.

Second, common-law powers must be exercised in conformity with statutory legislation. In this case, the main limitations stem from the right to private life under articles 8 European Convention on Human Rights and 7 European Union Charter of Fundamental Rights (EUCFR), the right data protection under article 8 EUCFR and the requirements of data protection law, found in the Data Protection Act 2018 and the EU General Data Protection Regulation (GDPR). A final potential legal basis is discussed in the following sub-section.

d. *Prerogative powers*

The final potential source of legal powers discussed here concerns the 'royal prerogative', i.e. the non-statutory powers of the Crown and its Ministers that were part of its extensive powers over its subjects in earlier times and have not been restricted by legislation.[603] The Law Commission concedes that 'the precise character and extent of the prerogative is a matter of debate'.[604] Endicott argues that a clear articulation of the principles that justify the power of

[600] Littlechild (n 268); the Law Commission (n 2) [9.1] agreeing.

[601] *Attorney General v De Keyser's Royal Hotel Ltd* (1920) AC 508; *R v Secretary of State for the Home Department, ex parte Northumbria Police Authority* (1989) QB 26.

[602] *R (Hooper) v Secretary of State for Work and Pensions* (2005) UKHL 29.

[603] Wade and Forsyth (n 591) 179.

[604] Law Commission (n 2) [1.47].

the executive in the British constitution has been undermined by a view of the prerogative as a 'stubborn stain' that would better be 'washed away'.[605]

Nonetheless, when it comes to research-data-sharing, the prerogative does not seem a good fit for two reasons. First, prerogative powers are usually *constitutional*, i.e. 'creating the basic institutions of the State and stipulating their powers and relationships with each other',[606] as opposed to administrative, in character. Bradley et al refer to powers relating to the legislature, the judicial system, foreign affairs, treaties, war and armed forces as prerogative examples.[607] Secondly, they are paradigmatically described as the legal powers which 'appertain to the Crown but not to its subjects',[608] which means that they are exclusive. Research data sharing, however, is much closer to an administrative operation that can also be exercised by natural persons. Unsurprisingly, prerogative powers were never brought up by participants to this study as one of the potential legal powers that data custodians could rely upon. The following section presents legal restrictions on legal powers to share data.

ii. Limits on Powers to Share Data for Research Purposes in UK Law

Legal powers to share data are far from unconstrained; the UK Supreme Court clarified in *Christian Institute* that they are 'conditional upon compliance' with the requirements of data protection and human rights law.[609] The two sources of limitations are discussed, in turn, in this section.

---

[605] T Endicott, 'Parliament and the Prerogative: From the Case of Proclamations to Miller' (2016) Judicial Power Project <http://judicialpowerproject.org.uk/wp-content/uploads/2016/12/Endicott-2016-Parliament-and-the-Prerogative.pdf>.

[606] House of Lords Constitution Committee, 'Constitution – First Report' (2000-01) <https://publications.parliament.uk/pa/ld200102/ldselect/ldconst/11/1101.htm> ch2.

[607] A W Bradley et al, *Constitutional and Administrative Law* (16th edn, Pearson 2015) 253 *et seq*.

[608] *R v Secretary of State for Health ex p C* (2000) 1 FLR 627 (Hale LJ); *Shrewsbury and Atcham Borough Council v Secretary of State for Communities and Local Government* (2008) EWCA Civ 148.

[609] *Christian Institute* (n 116) (Lady Hale et al) [58]. The common law duty of confidentiality may also apply to administrative data sharing for research purposes depending on the context, in addition to statutory data protection and human rights law.

a. *Limitations stemming from data protection law*

Data protection law determines whether personal information is processed 'fairly and lawfully'.[610] Administrative datasets, generally, contain information capable of identifying individual citizens. Intermediary bodies in the research data sharing space are often tasked with ensuring compliance with data protection law, e.g. the Information Commissioner's Office (ICO) which oversees the compliance of data sharing actors with the data protection safeguards prescribed in domestic legislation.[611] Is it, however, the case that research-data-sharing involves personal data in all cases?

The answer to this question is more elusive than it seems. For example, section 64(3) of the Digital Economy Act 2017 states that personal data must undergo de-identification, i.e. removal of any identifying information, before disclosure. The Administrative Data Research Network and other bodies like the SAIL DataBank have been using trusted third parties to de-identify and link the acquired datasets before allowing a researcher access.[612] Data shared under such standards, one might assert, are not personal data, since researchers are not provided with identifying information. Indeed, the House of Lords held in *Common Services Agency* that properly 'barnardised' (a method of de-identification) data are not personal data at the point of becoming available under a Freedom of Information Act 2000 request, as the recipient lacks the information allowing re-identification.[613] The ICO concur in their 'Anonymisation Code of Practice' that 'effective anonymisation' is possible through pseudonymisation (another de-identification method).[614]

Nevertheless, there are salient reasons for discussing data protection law limitations. De-identification is not a uniform legal requirement and there might be cases of data sharing which

[610] Section 85 of the Data Protection Act 2018.

[611] ICO (n 10).

[612] O Flanagan, 'Opening up admin data to research: an interview with Melanie Wright' <https://www.statslife.org.uk/features/1689-opening-up-administrative-data-to-research-an-interview-with-melanie-wright>; SAIL Databank, 'The Anonymisation Process' <saildatabank.com/saildata/data-privacy-security/#anonymisation-process>.

[613] *Common Services Agency v Scottish Information Commissioner* (2008) UKHL 47 (Lord Hope) [26].

[614] ICO, 'Anonymisation: managing data protection risk code of practice' <https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf> 21; Mourby et al (n 472) 222.

do not apply it. In that regard, public bodies in the UK seem to have been treating pseudonymised data as personal data. The National Health Service 'Confidentiality Code of Practice' stresses that pseudonymised information 'differs from anonymised information in that (…) it allows information about the same individual to be linked in a way that true anonymisation does not'.[615]

The main data protection legal instrument in the UK is currently the Data Protection Act (DPA) 2018 which has, practically, implemented the EU General Data Protection Regulation (GDPR) into domestic law. As mentioned before, administrative data pose the challenge of containing personal data which were originally collected at a different time and for a different purpose from research. Three of the data protection principles enshrined in section 85 of the DPA merit particularly careful consideration in the light of the 'fluid and norm-challenging nature of a social science research endeavour'[616]: lawful basis for processing, purpose limitation and data minimisation. For example, a data custodian like the National Health Service interested in providing access to researchers who want to analyse the impacts of a government policy on the health and well-being of citizens needs to ensure that these principles are satisfied. The legitimate aim of processing needs to be carefully defined and an explanation of why specific datasets are necessary to meet this aim needs to be offered.

With regard to lawful basis for processing, section 86 of the DPA implements article 6(1) GDPR which requires that any data processing activity have a lawful basis, providing an exhaustive list of potential legal bases. For present purposes, the relevant bases are consent [6(1)(a)] and necessary processing for the 'performance of a task carried out in the public interest or in the exercise of official authority vested in the controller' [6(1)(e)]. Recital 43 of the GDPR cautions against public authorities over-relying on consent for processing due to the imbalance of power between the controller and the data subject. While both the Information Commissioner's Office[617] and the Article 29 Working Party[618] clarify that the use of consent

[615] R Jay, *Data Protection Law and Practice* (Sweet & Maxwell 2012) 596.

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

[617] ICO, 'Consent' <ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/consent>.

[618] Article 29 Working Party, 'Guidelines on consent under Regulation 2016/679' (28 November 2017) <ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030> 6.

in public sector data processing is not totally excluded by the GDPR, there is an increased burden of proof for the public authority to demonstrate that the consent was freely given. This makes consent a less appealing basis in this context, both from a logistical and a legal perspective. Especially for large ministerial departments like the Department for Education or the Department for Work and Pensions, consent might not even be an option. Hence, invoking the 'public interest' / 'official authority' ground is the most appropriate basis for collecting and processing data for research purposes.

'Purpose limitation' comes into play when repurposing for research and evaluation use already collected data for other purposes, as is the case with administrative data. In the case of research data sharing, a tension seems to be created between Recital 50, article 5(1)(b) and article 6(4)(c) GDPR. More specifically, Recital 50 stipulates that further processing for a different purpose is allowed only 'where the processing is compatible with the purposes for which the personal data were initially collected', with the exception of, *inter alia,* 'scientific or historical research purposes' that should be regarded as 'compatible lawful processing operations'.[619] Article 5(1)(b) GDPR qualifies this presumption of compatibility by stressing that such further processing 'shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes.' This would seem to guarantee compatibility when e.g. the initial purpose is routine case management and the further processing takes place for research purposes, provided that the safeguards of article 89 (including data minimisation) are met. Considering that the European Data Protection Board have recently found the conditions for applying this presumption of compatibility 'horizontal and complex',[620] it is safe to assume that the data custodians will have to produce a tailor-made policy on the matter. For example, the Justice Data Lab clarify the precise variables they use to identify sentences and follow-up periods in measuring re-offending, explaining why these are necessary to ensure quality of the analysis.[621]

[619] Recital 50 EU GDPR.

[620] European Data Protection Board, 'Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (art. 70.1.b))' (23 January 2019) <edpb.europa.eu/sites/edpb/files/files/file1/edpb_opinionctrq_a_final_en.pdf> 8.

[621] Lyon et al (n 124) 232

Finally, yet importantly, research data sharing raises challenges from a data minimisation perspective, considering that data custodians may be inclined to retain-and-reuse datasets for research so as to minimise logistical costs related to data extraction.[622] Data minimisation, under section 90 of the DPA, however, dictates that personal data 'must be kept for no longer than is necessary for the purpose for which it is processed'. The research exemptions established by the GDPR and the DPA mitigate the rigour of these requirements to ensure that they do not 'prevent or seriously impair' the achievement of the purposes of processing,[623] i.e. in our case the facilitation of independent data-intensive research in the public interest.

That being said, the research exemptions are far from unconditional. For these exemptions to apply, research purposes must be the *sole* purpose of data processing, the Information Commissioner's Office having the power to determine whether this is indeed the case.[624] The processing should always be linked with the publication of academic material[625] and data custodians need to have a reasonable belief that such a publication would be in the public interest.[626] Provided that scientific research is the sole purpose of processing, it is not necessary to set a specific retention period and data can be kept for longer.[627] Therefore, the law allows significant scope for accommodating research data sharing without breaching the data protection principles. Human rights law is another significant source of limitations.

b. *Limitations stemming from human rights law*

Personal data sharing, especially when done on a massive scale, raises privacy issues even if it is only academic researchers that get access. Recent high-profile research data processing debacles in the UK, like Cambridge Analytica, illustrate that mishandling and manipulation of data are not the exclusive province of commercial entities.[628]

---

[622] Jackson (n 308).

[623] Data Protection Act 2018 schedule 2, part 6, paragraph 27.

[624] Data Protection Act 2018 s 174(3)(a).

[625] *Ibid* ss. 174(3) & 176(1).

[626] *Ibid* Schedule 2, Part 5, para 26 (2)(b).

[627] ICO, 'Principle (e): Storage limitation' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/>.

[628] P Lewis et al, 'Cambridge Analytica academic's work upset university colleagues' (2018) <theguardian.com/education/2018/mar/24/cambridge-analytica-academics-work-upset-university-colleagues>.

In UK law, breach of a fundamental right is a distinct ground of judicial review in accordance with section 4 of the Human Rights Act (HRA) 1998. The crucial provision here is the right to private life under article 8 of the European Convention on Human Rights (ECHR), which has been interpreted broadly to encompass numerous privacy interests including the 'fundamental right to protection of personal data' under article 8.[629] Considering that the UK Supreme Court held in *Christian Institute* that discussing the compatibility of data sharing legislation with article 8 ECHR suffices to determine compatibility with data protection,[630] it may be the case that human rights limitations will precede data protection considerations in litigation. Does article 8 ECHR, however, apply to all scenarios of research data sharing?

Privacy is a *qualified* right, i.e. one which lawfully accepts limitations when this is necessary in the pursuit of a legitimate aim. The ECHR adopts a two-staged enquiry: first, the Court considers whether there is an 'interference' with the right and, second, whether this interference is justified after balancing the relevant rights with other rights or interests. The discussion on de-identification and 'identifiability' is again relevant here. Lady Hale held in *South Lanarkshire* that it is difficult to see an interference with subjects' rights in cases where processing does not enable the data recipient or anyone else to identify the data subjects.[631]

Yet, human rights law limitations may be engaged without requiring a definitive legal solution of the identifiability conundrum. This is because interfering with a human right covers, essentially, all cases where the possibility of interferences *cannot be precluded,* rather than requiring their positive existence or a rigorous probabilistic assessment of their likelihood. In our context, Advocate General Cruz Villalón advocated in the CJEU case of *Digital Rights Ireland* for a broad notion of 'interference', encompassing all cases of storing and collection of personal data by public authorities regardless of the subsequent 'use made of them'.[632] The Advocate General emphasised that for such interference to arise, the data does *not* have to be 'personal data in the traditional sense of the term, relating to specific information concerning

[629] *Satakunnan Markkinapörssi Oy v Finland* (2018) 66 EHRR 8 at 74.

[630] *Christian Institute* (n 116) (Lady Hale et al) [104].

[631] *South Lanarkshire Council v Scottish Information Commissioner* (2013) UKSC 55 (Lady Hale) [26].

[632] *Digital Rights Ireland Ltd v Minister for Communications* (ECLI:EU:C:2014:238) Opinion of AG Cruz Villalón [69] citing *Amann v Switzerland* (2000) 30 EHRR 843 [65, 69 and 80] and *Leander v Sweden* (1987) 9 EHRR 433 [48].

the identity of individuals'. It can be data of a more aggregate character. This is often the case with the data held by prominent data custodians in the UK, with large ministerial departments often keeping datasets that do not include identifying variables, being interested in patterns rather than individuals.[633] Similarly, the UK Supreme Court held in *Christian Institute* that the statutory requirement of compliance with the Data Protection Act 1998 and the EU directive in section 26(11) of the Children and Young People (Scotland) Act (CYPA) 2014 did not preclude a potential interference with article 8 ECHR.[634]

At the second stage, the European Court of Human Rights adopts a proportionality test under article 8(2) ECHR.[635] Within this test, the 'public' or 'general' interest served by social research may be invoked as a legitimate objective.[636] Assuming, then, the suitability of data disclosures to attain this objective, there are two issues within proportionality,[637] i.e. whether data sharing powers are 'in accordance with the law' and whether they are 'necessary' for the achievement of the pursued aim.

Christian Institute shows how challenging both tests can be. The Supreme Court there struck down major provisions of part 4 of the CYPA 2014, which obliged or empowered public bodies to share personal data of young persons to a 'named person', when the information was 'likely to be relevant' to their wellbeing. There were two deficiencies in the Act. First, it was not 'in accordance with the law' because it was not sufficiently 'foreseeable and accessible' to prevent arbitrary interferences with privacy rights by public bodies.[638] Secondly, the Act did not establish safeguards that would enable 'the proportionality of an interference to be adequately

[633] Interview with participant nr. 10 (2018, Belfast).

[634] *Christian Institute* (n 116) (Lady Hale et al) [78-81].

[635] A8(2) ECHR: 'necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country (…) or for the protection of the rights and freedoms of others'.

[636] U Kilkelly, 'The right to respect for private and family life' (COE 2003) 30 on the European Court of Human Rights according wide margin of appreciation to member states in identifying a legitimate objective; O Lynskey, *Foundations of EU data protection law* (OUP 2015) 135 on the rights and interests balanced with data protection by the Court of Justice of the EU.

[637] *R (Daly) v Secretary of State for the Home Department* (2001) UKHL 26 (Lord Steyn) [27].

[638] *Christian Institute* (n 116) (Lady Hale et al) [79].

examined'.[639] In this respect, the Court noted that the ECHR compatibility of the Act could be rescued, had there been legally binding proportionality 'guidance to the information holder'.[640]

Research data sharing legislation is, arguably, a different case from CYPA, with research organisations being subject to a number of constraints. If we consider a few examples, section 253(1)(ca) of the Health and Social Care Act 2012 specifically requires that the Health and Social Care Information Centre (the custodian body) consider 'the privacy of recipients of health services and of adult social care' before disclosing data for research. The Digital Economy Act 2017, while not incorporating such an explicit balancing with privacy, reduces the likelihood of an interference through its requirements of de-identification [section 64(3)] and conformity with the UK Statistics Authority 'Code of Practice' [section 64(10)]. The latter is not merely advisory, but legally binding, since it regulates the researcher and project accreditation which is mandatory under sections 64(8)-(9) of the Act. Such accreditation is under the (current version of) the Code of Practice strictly dependent on adherence to its legal and ethical principles, which *inter alia* cover: data security and confidentiality, proportionality between the disclosure burdens and anticipated benefits, the construction of public interest, data retention, onward disclosure and transparency.[641] Hence, there are salient reasons to believe that existing research data sharing legislation satisfies, in principle, proportionality requirements in human rights law.

The analysis of the formal legal framework applicable to research data sharing in the UK demonstrates that the law sets the broader confines within which data sharing participants are enabled to create and apply data sharing frameworks. To what extent, however, does the interpretation of these legal powers and limits influence the actual construction of research data sharing institutions in the UK public sector?

C. Law's Limited Reach: Interview Data from the Four UK Nations

How do the participants in research data sharing interpret the interplay between multiple layers of legal powers and limitations, as established by the formal framework? As it will be argued

[639] *Ibid* [84].

[640] *Ibid* [101].

[641] UKSA (n 70) 3: 'All parties involved (…) are *required* to adhere to these principles' (emphasis added).

here, interview data collected in the four UK nations demonstrate the limited reliability of associations between particular 'legal' interpretations and the creation or implementation of organisational data sharing frameworks in practice. In some cases, it may seem like a particular type of 'legal' interpretation may influence a data custodian's appetite to create and apply a transparent data sharing framework. The previous chapter juxtaposed the Office for National Statistics (ONS) and the Department for Work and Pensions (DWP) as two antithetical examples with regard to their research data sharing approaches. The ONS has committed itself to comprehensive and publicly available conditions and strict timelines, whereas the DWP prefers to reserve much more discretion for its decision-makers and confines itself to informal treatment of data sharing requests.[642] The data sharing literature perception is that the DWP is taking a 'more conservative stance' and guarding 'its information closely',[643] whereas the ONS is regarded as an exemplary case of creativity and promotion of research data sharing.[644]

This picture was well-aligned with the way data custodians from the two bodies interpreted the law, even if their powers are broadly similar from a formal-legal perspective.[645] Custodians from the ONS seem to be interpreting the law 'as widely as we possibly could' and perceive of its constraints on research-data-sharing as *reasonable* and easy to satisfy in practice. An interviewee handling information governance at the ONS stresses that the law entrusts the Office with the responsibility to clarify 'reasonable' legal constraints:

> *'The legislation is constructed in that way to say you can share data (...) it's as easy as possible, with some reasonable constraints to protect confidentiality, which we are responsible to define.'*[646]

[642] *Supra* chapter 4, C.

[643] Jones and Elias (n 338) 72; Law Commission (n 2) [9.1].

[644] Ritchie (n 71) 3.

[645] For instance, the confidentiality obligations of the Department for Work and Pensions and the Office for National Statistics when disclosing personal information are governed by different legal instruments, i.e. sections 123 Social Security and Administration Act 1992 and 39 Statistics and Registration Services Act 2007, respectively. The two provisions, however, are substantially similar: they both prescribe a criminal offence of maximum two years for a civil servant who disclosed personal information held in relation to their functions when the explicit legal justifications ('made with lawful authority') do not apply.

[646] Interview with participant nr. 1 (2017, Titchfield).

On the other hand, custodians at the DWP seem to be much more concerned about the law's far-reaching limiting capacities when it comes to research data sharing. A policy-maker at the DWP explains that the Department is sceptical to use implied powers or its common-law powers to share data, preferring instead to rely on an 'explicit statutory gateway':

> '*We have eroded our common-law powers, having set a principle of using explicit legislation (...) the ground has been taken; we've said that this data can go to this organisation for this purpose (...) you might have something similar, but if it's not the same, we've set the standard.*'[647]

One might argue that the more permissive the legal interpretation a data custodian adopts, the more likely they are to establish and apply an organisational data sharing framework in practice.[648] There are reasons, however, to resist this interpretation. Qualitative data from the four UK nations illuminate why the relevance of 'legal' interpretations for the actual implementation of data sharing frameworks is limited in practice.

i. England: the letter of the law and the spirit of the 'public benefit'

To start with the example of England, interviewees who have worked for the Department for Education (DfE) explain that changes in the DfE's data sharing procedures were mostly associated with shifts in general beliefs within the organisation rather than legal considerations. An interviewee with a rich experience as a policy-maker in the DfE and other ministerial departments clarifies that even though a specific piece of legislation applies to DfE data sharing, i.e. the so called Prescribed Persons Act,[649] the actual mechanisms and procedures in place when they started their job there were 'quite weak actually'.[650] The same interviewee concedes that the law would be one of the factors considered when they 'tightened' their data sharing rules and procedures, yet not in a conclusive manner. They attribute this to the different ways of thinking between the legal team of the Department and the data sharing policy-makers:

[647] Interview with participant nr. 8 (2017, Sheffield).

[648] It shall be reminded that the aim of the present analysis is not to explain particular outcomes of data sharing decisions, but provide an in-depth understanding about the forces that structure and justify the process of decision-making, see *supra* chapter 2, A, i.

[649] Education (Individual Pupil Information) (Prescribed Persons) (England) Regulations 2009 (Amended 2013).

[650] Interview with participant nr. 30 (2020, virtual).

> *'There will be lots of conversations that the legal team would advise on what can or can't be done - they tend to take a black or white view to this. It's not always that way, sometimes it can be conditional upon certain things (…) The policy team, the customers, specify the policies that they are developing and how this will work.'*[651]

Another interviewee, who has both worked as a data analyst for the DfE and negotiated access to DfE data on behalf of an independent research organisation, reflects on this issue. They claim that changes in the legal framework were often not reflected in organisational data sharing procedures.[652] According to them, in spite of the Digital Economy Act (DEA) 2017 empowering the DfE to link their data with HM Revenue and Customs (HMRC) and Department for Work and Pensions data, the process of establishing a data sharing procedure to facilitate this has been 'quite a drawn out process'.[653]

An interviewee who has worked for the HMRC in a policy-making capacity corroborates this picture, stressing that, while law is one part of the picture, resource allocation is often a much more central consideration in designing a data sharing procedure. With all data sharing legal powers being 'discretionary' and permissive with regard to research purposes, resource management would often be the primary motivator behind revising the HMRC data sharing policy. As they recall, one of the key reasons behind establishing a uniform HMRC strategy on data sharing was that certain data shares were not cost-effective, even if lawful:

> *'Some areas would not share data, others would over-share - not illegally, but they would pour resource into something that was more trivial and could potentially be treated more strategically.'*[654]

Beyond large ministerial departments like the DfE and HMRC, the limited reach of law in determining the implementation of data sharing frameworks in practice is reflected in the

[651] *Ibid.*

[652] Interview with participant nr. 33 (2020, virtual).

[653] *Ibid.*

[654] Interview with participant nr. 32 (2020, Oxford).

earlier practice of the Office for National Statistics (ONS), before they established their Virtual Microdata Laboratory (VML). Before access became formalised, external researchers who were interested in ONS data had to sign the so-called '1-pound contracts', under which they were nominally appointed by the Government as employees.[655] An interviewee who has worked for the ONS as a policy-maker and data analyst vividly recalls how the Office's legal team conceded that this informal practice was not consistent with the law:

> *'I went and met the legal people (…) And they, basically, looked at this and said, well, the 1 pound contracts are not really within the spirit or the letter of the law, but there isn't really any mechanism, because this data is only available for government departments.'*[656]

In the interviewee's own opinion, to argue that there was a 'solid legal foundation' for the 1-pound contracts is 'pretty much nonsense'.[657] Yet, they claim that setting up this data sharing procedure was 'definitely within the spirit of working for the public good', which 'trumps the spirit of the law'.[658] The same interviewee goes as far as saying that, in their view, the law cannot be seen as 'relevant' when designing a data sharing framework:

> *'I rarely talk about the law, but I always talk about the ethics. For me, the law is not relevant, law is a thing you do. We are going to stick within the law, that's taken, that's like saying we are going to obey the law of gravity! We don't really have much choice on this, we will follow the law.'*[659]

This is consistent with the opinion of another interviewee, currently working for the ONS as a policy-maker. When reflecting on the challenges around developing a comprehensive data re-use policy for research purposes, they claim that this is '*not a legal issue'* but a matter of '*just getting our ducks in a row to be able to do it'*.[660] Certainly, understanding legal powers and limitations is always part of designing a data sharing procedure, but these accounts indicate

[655] Ritchie (n 71).

[656] Interview with participant nr. 42 (2020, virtual).

[657] *Ibid.*

[658] *Ibid.*

[659] *Ibid.*

[660] Interview with participant nr. 4 (2017, Titchfield).

that data custodians in England do not see the law as the dominant source of research data sharing institutions. In other words, legal interpretations do not hold the keys to explaining why a particular organisation may have implemented its data sharing framework in a specific manner.

ii. Northern Ireland: the law barks, but does not bite

The picture in Northern Ireland is substantially similar to England, but also, in a way, distinctive. Northern Irish data participants seem to be combining *both* a restrictive perception of law as robust and strict *and* an appetite to establish transparent and clear procedures for research data sharing. In that sense, a particular type of legal interpretations is not necessarily associated with a reluctance to establish transparent data sharing procedures in practice, as one might expect.

This is reflected in the views of two interviewees working in an information governance and policy-making capacity for the Business Services Organisation (BSO) in Belfast, one of the pioneer research data sharing bodies in the Northern Irish context. BSO, part of the Department for Health and Social Care in Northern Ireland, is a data custodian that has developed a transparent organisational data sharing procedure – the Honest Broker Service – (partly) to provide researchers with access to its health and social care data.[661]

While working for one of the most transparent and researcher-friendly custodians in Northern Ireland, their perceptions of law could be interpreted as rather restrictive. They clarify that the stage of assessing the legality of the proposed data share by the 'personal data guardian' of the custodian (usually the head of their legal services) is an important one. When they are asked to provide their overall impressions of the legal framework, an interviewee who works in information governance admits that the same legal regulations can be interpreted differently, depending on an organisation's '*risk appetite*'.[662] The BSO, they claim, demonstrate a quite low level of risk appetite, due to the sensitive nature of health and social care data. Their policy-

[661] Health and Social Care Northern Ireland (n 248): 'The HBS enables access to anonymised, aggregated and in some cases pseudonymised health and social care data to the Department of Health, Health and Social Care organisations and for anonymised data for health and social care related research'.

[662] Interview with participant nr. 11 (2018, Belfast).

making colleague concurs with and expands on that thought, highlighting the significance of strict legal limits on data sharing:

> *'I think we need to have hoops to jump through. We have very, very sensitive data here. My data is part of that. I want that to be protected and the risk of identifying me to be as small as possible. Obviously, sometimes I'm wearing my researcher hat. But I understand that there is a spectrum of risk from very low to very high and we want it to be very low, but we don't want to prevent research from happening at all.'*[663]

Another interviewee, working for the Northern Irish Department for Employment and Learning as a data analyst, finds the legal landscape '*not permissive enough in terms of data sharing'*,[664] especially before the enactment of the UK Digital Economy Act 2017. A similarly restrictive view is expressed by another individual working for the same data custodian:

> *Legal gateways must be clearly defined and established prior to data linkage. (…) Importantly, once the aims of the research have been defined the data cannot be used for alternative purposes. This removes the risk of data sources being linked for undefined purposes or simple data 'fishing expeditions' which would be in breach of data protection legislation.* [Statement on file with the author][665]

With a doctrinal-legal analysis indicating that this individual is overstating the rigour of data protection law constraints on research data sharing,[666] their perceptions of law can be convincingly understood as falling on the 'restrictive' end of the spectrum. Yet, as it will be recalled from the previous chapters, research data sharing has been much more streamlined and

[663] Interview with participant nr. 12 (2018, Belfast).

[664] Interview with participant nr. 16 (2018, phone).

[665] 'Linking Administrative Data: recent experience within the Department for the Economy' (on file with the author) – note that the Department for the Economy (NI) has succeeded the Department for Employment and Learning (NI). This individual was approached for an interview, but preferred to provide the author with a (comprehensive) written statement of their view on the matter instead.

[666] Bell et al (n 121) 46.

transparent in Northern Ireland compared to England.[667] An interviewee who works in an information governance capacity for the Northern Irish Statistics Authority clarifies that this progress in formalising and streamlining research data sharing has not come at the price of downplaying the robustness of legal safeguards. When asked about the crucial stages of a data access request, they clarify that identifying '*the legal power to both receive and process admin data for research*' is a primary consideration.[668] Mirroring the response of the interviewees working for the Business Services Organisation,[669] they clarify that their risk appetite when it comes to the legality of data sharing is really low. In their words, it is of '*utmost importance*' that a department's solicitors are happy with the existence of an '*explicit legal gateway for both the reception and the processing of the data*'. They see an attempt to go beyond the letter of such a gateway as '*trying to fudge things or find another way or make another way*'.[670]

The limited reach of law is corroborated in cases where an explicitly restrictive perception of legal powers and limitations is not expressed. Another interviewee, who works for Queen's University Belfast as a researcher and is affiliated with the Administrative Data Research Centre-Northern Ireland, attributes organisational responses to research data sharing in Northern Ireland to 'culture' rather than changes in the law.[671] While they recognise that the 'legalities' around the Digital Economy Act 2017 were an issue to be carefully considered by data custodians, they see the establishment of data sharing procedures as largely unrelated to legal requirements:

> *'This is about conversations, relationships (…) They know me, they know what type of questions I am going to ask - and when you get the right people around the table, and we get a collective decision, and then we have progress. So, actually, it depends on the culture.'*[672]

The same interviewee recalls that law was only a small part of resolving a major governance challenge within the Administrative Data Research Centre-Northern Ireland, i.e. establishing a

[667] Research Project Management (n 29); *supra* chapter 4, C.

[668] Interview with participant nr. 9 (2018, Belfast).

[669] Interview with participants nr. 11 and 12 (2018, Belfast).

[670] Interview with participant nr. 9 (2018, Belfast).

[671] Interview with participant nr. 10 (2018, Belfast).

[672] *Ibid.*

set of requirements for appropriate linkage of consented survey data to non-consented survey data or administrative data.[673] Also reflecting on the shortcomings of National Health Service England's care.data initiative to gain social acceptability,[674] they stress that establishing rapport with the responsible Research Ethics Committee in Northern Ireland was much more central in developing their own position on the matter.[675] Even if the legal framework allows such a data linkage under the lawful basis of 'public interest' data processing, ethical and social acceptability are crucial for the long-term viability of a data sharing process like this. In the interviewee's own words:

> *'The premise we have reached so far is that legally we can - legally we can link it (...) but socially do you want to? How many times would someone expect to answer a lot of questions only to find that this data has been linked behind their backs without their consent? Socially, how long you can do that for?'*[676]

The presented here cases of Northern Irish data participants demonstrate that particular legal interpretations are, generally, not determinative of the establishment of a data sharing framework. Interpreting the law as a barrier does not suffice to steer a data custodian towards inertia and cautiousness, considering the transparent and research-facilitative frameworks adopted by Northern Irish data custodians like the BSO or NISRA. As the following chapters will explain, other, more salient, factors are associated with the establishment of such data sharing frameworks.

iii. Wales: the law as a post-hoc justification

The Welsh data sharing landscape presents us with a paradox similar to the Northern Irish landscape. On the one hand, conservative and restrictive interpretations of legal powers and limits on sharing data to external researchers. On the other, the existence of a set of sophisticated, research-facilitative and transparent data sharing frameworks, such as the one

[673] Robinson et al (n 453).

[674] Carter et al (n 206).

[675] Interview with participant nr. 10 (2018, Belfast).

[676] *Ibid.*

developed by the SAIL DataBank which includes data from the majority of health data providers in Wales.[677]

Like Northern Irish data custodians, Welsh data custodians seem to be conservative in their interpretation of what the law empowers them to do. An interviewee, working in a policy-making capacity for the Welsh Department of Health, stresses that the law is not facilitative for research data sharing purposes:

> *'It is too restrictive, (…) in the way that it's complex. There are so many different bits that determine what and how you can do. (…) If there was one single piece of statute (…), it would be easier to proceed within the current framework. But because there's lots of different bits, including common law that apply to this space. It makes it more complex than it needs to be.'*[678]

Another interviewee, performing information governance duties for the National Health Service Wales Informatics Service (NWIS), which is effectively a data custodian for a number of National Health Service Health Boards and Trusts in Wales, concurs. They point out that the law imposes many '*formal requirements*' for organisations participating in research data sharing, causing '*a little bit of confusion*' due to the complex interplay between the recently introduced EU General Data Protection Regulation and domestic data protection law.[679] Another interviewee, working in a policy-making capacity for the Administrative Data Research Centre-Wales, concedes that the inexistence of an explicit legal gateway, particularly with regard to acquiring central government data, remains one of the most persistent challenges for the Network.[680] This seems to be at odds with the assumption, made in the literature,[681] that restrictive legal interpretations result in organisational inertia and cautiousness with regard to research data sharing. How should we, then, explain the tendency of Welsh data custodians to create and implement research-facilitative and transparent frameworks in spite of being conservative about their legal powers?

[677] Ford et al (n 221).

[678] Interview with participant nr. 24 (2019, phone).

[679] Interview with participant nr. 23 (2019, Cardiff).

[680] Interview with participant nr. 21 (2019, Swansea).

[681] Law Commission (n 2); Administrative Data Taskforce (n 23); Littlechild (n 268).

An explanation of the apparent paradox comes in the eloquent explanation of the interviewee working in an information governance capacity for NWIS. As they explain, organisational decisions to establish a data sharing framework are often shaped by different considerations, with law considered only *after the fact* to ensure that the established framework is compliant with legal requirements. In other words, law often operates as a *post-hoc* justification:

> *'What tends to happen is you make the decision, you then look at the Regulations and see whether you can back that up with what the Regulation says.'*[682]

This is why the same interviewee claims that the 'sky has not fallen in' after the introduction of the General Data Protection Regulation. In their view, legal requirements are 'essentially the same but put into slightly different context potentially'. Hence, '*culturally changing organisations*' to reassure them that disclosing personal information is the right thing to do is much more crucial than the legal barriers:

> *'It is not just the law we are complying with, this is also about culturally changing organisations in order to be assured that the sharing of personal information is the right thing to do. Sometimes this can be a barrier more than the technical or the legal aspects.'*[683]

This is consistent with broader reflection of Welsh data sharing participants on the limited reach of legal interpretations in determining organisational solutions to research data sharing. Two interviewees working for the Knowledge and Analytical Services at the Welsh Government agree that law '*has done what it can*' and more law cannot help with streamlining research data sharing procedures.[684] They explain that since the Digital Economy Act 2017 laid out some broad principles on data sharing requirements, other aspects like organisational incentives are the '*biggest issue*' in this context.[685] Another interviewee, working in a policy-

[682] Interview with participant nr. 23 (2019, Cardiff).

[683] *Ibid.*

[684] Interview with participant nr. 25 (2019, Cardiff).

[685] Interview with participant nr. 26 (2019, Cardiff).

making capacity for the Welsh Department of Health expresses a similar view: if you want to see changes and improvements in data sharing frameworks, do not expect it from the law. Since legislative changes often take 'at least three years', seeking to establish consistent interpretations among organisations on the ground is crucial to establishing mutually compatible data sharing frameworks.[686]

An interesting, complementary explanation of the apparent paradox is offered by an interviewee who works for the SAIL DataBank as a data analyst. As they explain, seeing the law as a barrier to establishing a transparent data sharing procedure is often caused by the distribution of labour within data custodian bodies. Legal teams within data custodian bodies will normally express the organisational position on research data sharing and the duty of such teams is to make sure the organisation avoids 'being sued', much like 'a technical person might be worried about exposing their system externally'.[687] As they explain, however, resource-related limitations in the civil service may be more salient barriers than the law:

> *'A lot of the time, people interpret law as barriers, but perhaps they are not (…) You are dealing, sometimes, with statisticians, sometimes with technical persons, sometime with an administrator who does data entry. They've got day jobs to do. If there was a single person's job to facilitate data access, that would be great, but budgets are squeezed as well.* '[688]

In their view, data custodians are happy to engage in 'constructive conversation' about their interpretation of legal requirements, as long as data sharing collaborators like the SAIL DataBank can convince them about their credibility and the overall merits of streamlining research data sharing.[689] In a similar vein, an interviewee working for the Confidentiality Advisory Group in England and Wales, a body responsible for approving requests for use of National Health Service data without consent,[690] clarifies that law is only a part of their focus.

---

[686] Interview with participant nr. 24 (2019, phone).

[687] Interview with participant nr. 22 (2019, Swansea).

[688] *Ibid.*

[689] *Ibid.*

[690] Health Research Authority, 'Confidentiality Advisory Group' <https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/confidentiality-advisory-group/>.

Although the Group's remit is to work within the boundaries set by the law,[691] the 'confidence of the public in the use of their data' is a key guiding principle for the interpretation of the Group's powers and duties.[692] This is also reflected in the fact that the current composition of the Group does not include any members who are 'purely lawyers' by background.[693]

Findings from England, Northern Ireland and Wales have indicated how differently expressed or justified narratives about research data sharing converge on the same conclusion: legal interpretations of data sharing powers and limits, either facilitative or restrictive, are not necessarily associated with the actual establishment and the character or organisational data sharing frameworks in practice. Does Scotland present us with a different picture?

iv. Scotland: it's not the law, it's the lawyers

The answer to this question is not straightforward. Scottish data sharing participants seem to be thinking of law as a more central consideration in the context of developing and implementing research data sharing frameworks. When pressed to clarify their thoughts, however, they quickly associate difficulties with legal *staff*, i.e. legal teams responsible for data sharing, rather than substantive legal hurdles to data sharing. Ultimately, it becomes clear that, to Scottish data sharing participants, it's not the law, but the lawyers that often block or enable the implementation of data sharing procedures, often for reasons unrelated to the law itself, such as staff changes or shortage of hands.

In a similar fashion to many interviewees from the other three UK countries, Scottish participants agree that the law applicable to research data sharing is complicated and creates difficulties in implementing a data sharing framework. This is not to say that participants necessarily see this as unreasonable, considering that data held by custodians is often sensitive and robust safeguards should be applied to minimise risks of breaches and reidentification. For example, an interviewee with both a clinical and a research background at the NHS finds the

[691] Section 251 of the NHS Act 2006 in conjunction with Regulation 2, 3(4) or 5 of the Health Service (Control of Patient Information) Regulations 2002.

[692] Interview with participant nr. 27 (2019, phone).

[693] *Ibid.*

legal framework 'mostly reasonable, sensible and fair',[694] with only a few exceptions.[695] Another interviewee, working as a researcher at the University of Edinburgh, agrees that robust legal safeguards are needed:

> *'Because I understand the constraints of it and the potential implications of using people's data or using people's biological samples without consent. I'm fully aware of that balance (...) I think that it is right to have those systems in place.'*[696]

Nonetheless, reasonable or not, some participants elaborate on how the law creates complications from the perspective of a requesting researcher or an intermediary body which seeks to facilitate research data sharing. An interviewee who works as an administrative data researcher at the University of Edinburgh explains that 'because the law is so complicated' and many data custodians do 'not really have a handle on it', researchers are often faced with the task of justifying the legality of a proposed data share.[697] They express scepticism about the extent to which the investment of time and resources needed for this justification is legitimately required from researchers:

> *'That becomes a big challenge - how much of it is our responsibility as researchers to say ''actually, no, you are ok, look at this'' - and how much is up to them is another question...'*[698]

An interviewee who works as a policy-maker for the Scottish Centre for Administrative Data Research, concurs that the law governing research data sharing is 'very cumbersome' and that 'academics are getting fed up with it'.[699] Their reflection of the rules is that:

---

[694] Interview with participant nr. 39 (2020, virtual).

[695] *Ibid:* 'My view is that this person is dead, the data should enter the common domain, and it should be stripped of certain things e.g. that they had a particular disease like syphilis or that they had gone bankrupt or something like that. Once a person is dead, you should be able to publish individual details, with some limits'.

[696] Interview with participant nr. 41 (2020, virtual).

[697] Interview with participant nr. 29 (2020, phone).

[698] *Ibid.*

[699] Interview with participant nr. 31 (2020, virtual).

> *'They are (...) really really arduous, many many ends of pages (...) It's quite a complex system - certainly, I think (...) It takes me a year to get permission and I have to sign away my life!'*[700]

The same interviewee stresses that the Digital Economy Act (DEA) 2017 has not lived up to the expectations surrounding it as the reform that will simplify research data sharing, at least up to the present moment. Nonetheless, they express some optimism about the future potential of the DEA to streamline data sharing procedures.[701] Another interviewee, working as an administrative data researcher at Edinburgh Napier University, is more sceptical. They feel that 'it is probably limited what the law can enable'.[702] To them, it is more about people *understanding* better what the law requires from them and less about a 'new law that would magically solve everything'.[703]

'Understanding' is a crucial term here and it brings the analysis to a crucial question: whose legal 'understanding' prevails and shapes the way in which organisational actors understand what the law requires from them or enables them to do in the context of research data sharing? It is the understanding of the legal teams' that are working for organisations, primarily for data custodians, who are the ones responsible for handling the legalities of data sharing,[704] but also in some cases for research organisations who want to ensure that 'the data we will receive is really what we should receive'.[705] An interviewee working for the Electronic Data Research and Innovation Service, an intermediary body in Scotland, expresses the view that 'the law has never hindered anything'.[706] Yet, there are still many conversations and 'different perspectives' on interpreting the law expressed by their legal colleagues who are to be relied upon for such matters:

---

[700] *Ibid.*

[701] *Ibid*: 'I imagine, in time, the DEA will prove to be the legal gateway that is most useful, but it's taking longer than we expected'.

[702] Interview with participant nr. 34 (2020, virtual).

[703] *Ibid.*

[704] Law Commission (n 2).

[705] Interview with participant nr. 41 (2020, virtual).

[706] Interview with participant nr. 36 (2020, virtual).

> *'We would always go to ask the CG or the DPO about that, since we're not experts on it. We would seek advice from our colleagues, they are more in the know than we are.'*[707]

An interviewee working as an administrative data researcher at Edinburgh Napier University agrees that the standstill generated in the application of data sharing frameworks around the time that the EU General Data Protection Regulation (GDPR) entered into force had to do with 'people misunderstanding GDPR quite a lot'.[708] They associate cautiousness in designing data sharing processes around that time with an increasing awareness that legal teams need to reassess compliance risks for organisations like data custodians:

> *'GDPR does not stop you from giving me the data, you can do this, even if you need the safeguards. I think it's understanding the law, rather than the law itself.'*[709]

Why is it, then, so difficult for legal understandings to converge on what is required for a robust, yet consistent and research-facilitative data sharing framework? The answer seemingly lies beyond the substance of legal norms and closer to the agonising realities of an austerity-struck, resource-limited public administration. An interviewee, who works as an administrative data researcher at the University of Glasgow and has been negotiating access to Department for Work and Pensions (DWP) data, reflects on how understanding what the law requires is a *long* process:

> *'With DWP, I think, it is still not clear about the legal basis or how they need to proceed. (...) They are still working on that, it's something that was a bit puzzling. We've been doing this for years, and now issues are created. I think the transition of the new law is still undergoing, there are still things that need to be clarified.'*[710]

---

[707] *Ibid.*

[708] Interview with participant nr. 31 (2020, virtual).

[709] *Ibid.*

[710] Interview with participant nr. 38 (2020, virtual).

It is, then, of little assistance to the aim of establishing consistent research data sharing frameworks that lawyers, as many other officials, are often 'leaving their positions'[711] and staff with information governance responsibilities keep getting shifted around to ensure that the public sector does 'more with less'.[712] An interviewee with a rich experience in policy-making positions within various ministerial departments describes this way of thinking:

> *'Pretty much all the time that I was working in DWP and DfE, I had a headcount challenge - in most organisations, you have some kind of financial target, you have to deliver so much business. Mine was kind of the reverse, it was about reducing my headcount - my full-time headcount. They would expect me to save 10%. And obviously do more with less staff.'*[713]

This lack of consistency and stability within legal teams, as well as the relative shortage of staff, resulted in a standstill when data custodians had to amend their data sharing frameworks in light of the GDPR. As an interviewee who works as an administrative data researcher at the University of Edinburgh recalls, many organisations would go as far as 'shutting down for a couple of weeks to figure out if they are compliant' with data sharing requests.[714] Another interviewee, working in a policy-making capacity for the Scottish Centre for Administrative Data Research, attributes this to fears around 'the issue of responsibility'.[715] In the absence of long-standing, converging cross-organisational standards around research data sharing, legal officials working for data custodians do not want to innovate in authorising disclosures:

> *'I think it's about the law as well, different interpretations of the legal requirements and descriptions (...) You don't want to be the person who is responsible for a data breach.'*[716]

Hence, according to Scottish research data sharing participants, legal interpretations do influence, to some extent, the establishment and implementation of data sharing frameworks.

[711] Interview with participant nr. 29 (2020, phone).

[712] Ferlie (n 304) 6.

[713] Interview with participant nr. 30 (2020, virtual).

[714] Interview with participant nr. 29 (2020, phone).

[715] Interview with participant nr. 31 (2020, virtual).

[716] *Ibid.*

They do so, however, for reasons that have less to do with the law itself and more to do with organisational limitations in investing into long-standing and properly staffed legal teams responsible for research data sharing. How are we to understand and explain such limitations? More broadly, in light of the limited reach of the 'law' in all four UK countries with regard to the formation of research data sharing institutions, where should we look at to identify the 'rules of the game' that govern research data sharing?

D. From Law's Limited Reach to the Salience of a Research Data Sharing Culture

'Legal' interpretations have been first defined and, then, their relevance to the regulation of research data sharing decisions was assessed in this chapter. Proceeding from an interpretive definition of 'legal' interpretations with the formal legal framework as its starting point, this chapter has drawn on interview data collected in the four UK nations to assess the relevance of law to the construction of organisational research data sharing frameworks.

The analysis advanced in this chapter indicates that, although lived experiences of research data sharing are not necessarily uniform in the four UK countries, they significantly converge as to the law's limited reach in determining the actual establishment and implementation of data sharing frameworks. In the different countries, participants may substitute the 'spirit of the law' with the 'spirit of the public benefit' (England), pay lip service to the law's strictness (Northern Ireland), use the law as a post-hoc justification for a particular data sharing procedure (Wales) or highlight the difficulties with legal teams (Scotland). In all cases, most of them would feel that legal interpretations are *not* the strongest factor behind the establishment of a particular organisational approach to research data sharing.

What are, then, these other stronger factors? If, in other words, the law is not the most salient source of research data sharing institutions, *what* is? A few alternative factors have been implied in the preceding analysis, such as culture, resources and relationships. Nonetheless, the present argument remains a *negative* one for the time being: it makes a case for what is *not* primarily shaping research data sharing regulation, instead of proving what *does*. It will be for the following two thesis chapters to make a case for how other institutions are more salient than the law in filtering access to administrative data for social-scientific research purposes.

6. THE SALIENCE OF A 'RESEARCH DATA SHARING CULTURE' FOR THE REGULATION OF RESEARCH DATA SHARING DECISIONS

> *'It doesn't matter if you have the best security systems of the world, if you have the best researchers, all the institutional guarantees in the world, well-written projects. If the culture of that Department or the culture of the person responsible is to not share the data or to be risk-averse, or for whatever reason the culture is to say no, instead of saying yes, no system or hypothesis in the world is going to make a difference'* [Interviewee 10]

> '*From first-hand experience of being in a meeting, asking them to provide data for specific purposes, being told that they could not see where there was any value for them. I think it's very much in the front of their mind in terms of ''what's in it for us'''* [Interviewee 31]

> '*We can see that it can be of interest in a general sense, but truly in the public interest as would the tax payers expect us to be putting our efforts into this, that's a different thing. We actually describe this as the 'so what?' questions. We can sometimes see requests for data sharing for something that demonstrates a very interesting methodological question, but of no practical use whatsoever and we have to legitimately ask the question: is that in the public interest?'* [Interviewee 8]

If the law is not the most salient driving force in the regulation of research data sharing decisions, what is? As this chapter will demonstrate, the answer is: a collective 'research data sharing culture', permeating the decision-making process and steering data custodians and other research data sharing participants towards committing themselves to conformity with certain rules and principles. This 'research data sharing culture' effectively shapes the formation of the 'rules of the game' in the research data sharing space, resulting in the prevalence of particular research data sharing institutions,[717] e.g. organisational policies that

[717] *Supra* chapter 2, C on the broad definition of institutions adopted in this study including such types of institutions as uncodified conventions and normative beliefs about the appropriateness of research data sharing.

allow data access by reference to such specific criteria as the relevance of a request to the data custodian's policy priorities.

More specifically, drawing on this study's empirical findings, this chapter unpacks the two core constituent elements of this culture: 'perceptions of administrative data ownership' and 'assumptions about the appropriateness of policy-relevant research'. These unwritten, uncodified, cultural forces, it is argued, predominantly shape the creation and application of research data sharing frameworks in the UK public sector. 'Perceptions of administrative data ownership' refer to the formation of collectively held attitudes towards *whose* data is being shared for research purposes. 'Assumptions about the appropriateness of policy-relevant research' relate to the ways in which data custodians evaluate the alignment of a proposed research request with their organisational policy priorities and their capacity to deliver them. This chapter will rely on interview data from the four UK countries, involving fourty-two (42) data sharing participants working in a variety of roles, i.e. as *data analysts* technically managing the datasets, *information governance officials* responsible for the compliance of data sharing with legal requirements and *policy-makers* who are responsible for higher-level decisions of data sharing policy, including the ultimate approval of data access requests for custodian bodies. The data also involve representatives from a variety of organisations, from large ministerial departments like the Department for Work and Pensions and the Department for Education, to academic research groups like the Administrative Data Research Centre – Scotland and bodies responsible for overseeing privacy safeguards in data sharing like the Confidentiality Advisory Group. These data illustrate the salience of the previously mentioned conventional, non-articulated, non-codified normative perceptions and assumptions in filtering access to administrative data for research purposes.

The 'rules of the game' in the research data sharing space are not uniform across the UK public sector,[718] yet qualitative data analysis yields specific conclusions about key divisions in organisational approaches to research data sharing. Particular views as to the ownership of administrative data, or as to the desired alignment between independent research and a custodian's policy aims, are often shared collectively. Such views are associated, as it will be shown, with the creation and implementation of particular types of data sharing frameworks.

[718] *Infra* chapter 8, C on the potential limitations of the present study in sketching the picture of different 'rules of the game' across the UK public sector.

At times, the result is the establishment of transparent data sharing procedures, whereas, at other times, transparency gives way to entitlement and discretion, resulting in independent researchers and intermediary bodies being kept in the dark by data custodians. In other cases, these views result in the establishment of particular powers of data custodians, such as their right to intervene and co-design the research questions that will be addressed by the independent researchers. The two constituent forces of the 'research data sharing culture', as well as their influence on the creation and implementation of research data sharing frameworks in the UK public sector, are analysed in turn.

A. Perceptions of Administrative Data 'Ownership': Sharing Whose Data?

Some actors populating the research data sharing space treat administrative data as a 'public good', much like other types of public infrastructure like roads or parks,[719] that should be used to produce research in the public interest. On the other hand, some other actors proceed from the opposite premise, i.e. they treat administrative data as belonging to the data custodians that are responsible for collecting and curating it. This division is clearly reflected in the divergent practices in respect of creating implementing research data sharing frameworks: custodians who view administrative data as a public good are more likely to establish specific procedures and criteria, becoming accountable to the research community and the broader public with regard to their data sharing decisions. Custodians who treat data as their ownership are more likely to resist accountability, thus avoiding to commit themselves to particular data sharing criteria and processes.

i. Data as a 'public good': enabling research and scrutinizing the government

Contrary to the relatively obscure legal position on the ownership of public-sector administrative data,[720] research data sharing actors hold strong opinions on who owns administrative datasets and, thus, who should profit from their use in research. While data custodians are, in practice, responsible for green-lighting any possible use of the data, including data sharing for research, there is a growing aspiration among some inhabitants of this space to ensure that administrative data are to be treated as a common resource. These actors have

[719] House of Commons Science and Technology Committee (n 208).

[720] *Supra* chapter 5, B.

sought, time and again, to negotiate with other inhabitants of the space, with a view to making their view more influential and widespread in the UK public sector.

Proponents of this 'common resource' view concede that data custodians curate the data and perform stewardship functions, including deciding when it is appropriate to allow access by researchers. Nonetheless, in their view, the data should be, in principle, available for use when a research project serves the public interest and the onus to prove that a disclosure is not desirable for some reason should rest with the data custodian. As the following discussion will illustrate, there are two main ideals behind viewing administrative data as a public good. First, since such data belongs to the public as a whole, it should be used in public interest research that will improve our collective knowledge about pressing social and economic challenges. Second, since such data belongs to the public as a whole, it should be used to scrutinize the validity of government decisions, as well as the evidence on the basis of which these decisions were taken.[721] A researcher working with administrative datasets in Scotland associates this view with communitarian political ideals, originally expressed with regard to public-sector data in Scandinavian countries:

> *'Denmark and Norway are much more oriented around the idea of social communities. The whole thing about the rights of the individual versus the rights of the general public, and getting that balance right. In those countries, more social-minded countries, the balance is geared towards the public.'*[722]

The Office for National Statistics (ONS) in England has been the torchbearer of this aspiration in the UK. An interviewee, with significant background as both a data analyst and a policy-maker for the ONS in England, explains that it is important to think of data availability for research purposes as the *default,* not the aim of a data sharing policy. For them, convincing their colleagues at the Office about this through constant bargaining has been one of their key 'achievements'. After many internal discussions, they recall that they managed to convince the information asset owners within the ONS that the data is 'fair game for researchers (…) if you could come up with a valid reason for using it'.[723] In their view, this was a significant departure

[721] This echoes the 'accountability' narrative behind research data sharing, as sketched in *supra* chapter 4, B.

[722] Interview with participant nr. 41 (2020, virtual).

[723] Interview with participant nr. 42 (2020, virtual).

from the previous position within the Office, under which researchers had to 'ask the various data holders and then get sign-off from the information asset owners'.[724] They recount vividly the process of negotiating this change of policy with their colleagues who were responsible for managing the datasets:

> *'Going around in the department, to the Information Asset Owners (...) and saying, basically, you have all of this data under your control - how about this, we work on the assumption that all the data is available for research use. Rather than me having to justify use of it, we want you to justify why you can't share it.'*[725]

The above-described incident is, arguably, one of the moments which marked the cultivation of ideas around achieving 'transparency' and 'accountability' through research data sharing within the ONS.[726] Staying in the same example, an information governance official working for the ONS places 'sharing as much data as possible' and 'enabling research' high in the Office's organisational priorities.[727] The Office's commitment to enable administrative data research in the public interest does not waver even if official Government policies are likely to be scrutinised by the research. As the same official explains, there are 'public benefits' in proving that government statistics are incorrect or that better measures of such phenomena as inflation can be devised.[728] Another interviewee, working for the ONS as a data analyst, makes an even stronger claim: it is *because* administrative data research is used to scrutinise government policies and their delivery that the Office is keen to enable such research. In their view, this is 'very important for democratic debate, for the public to actually know how effective the government is'.[729] This is why it is of utmost importance that any outcomes from the research 'should be published and made available'.[730]

---

[724] *Ibid.*

[725] *Ibid.*

[726] *Supra* chapter 4, A-B has elaborated on the pioneer role of the Office for National Statistics in invoking these narratives to widen access to data for research purposes.

[727] Interview with participant nr. 1 (2017, Titchfield).

[728] *Ibid.*

[729] Interview with participant nr. 3 (2017, Titchfield).

[730] *Ibid.*

Even before the ONS pioneering initiatives, administrative data researchers have been going at great lengths to support the broader adoption of this understanding of administrative data as a shared resource. In some cases, convincing data custodians about the merits of this view may necessitate the adoption of a more conflict-oriented approach from researchers. An interviewee, working as a data analyst for an independent research organisation and with significant experience in negotiating access to Department for Education (DfE) data, admits that they often aim to 'be hard and vocal' in their critique of education policies.[731] Interestingly, they explain that they would often find an unlikely ally in this effort in the face of DfE officials who were keen on seeing external scrutiny of their policies:

> *'Actually, many from the department may be welcoming criticism, because it's easier for us to do, rather than carry it out as an internal project.'*[732]

Another interviewee, with a similar background, explains that custodians might prefer external use of their data in scrutinising their policies because 'neutral' and 'independent' criticisms of government work often have 'more currency' than internal research.[733] An interviewee, working as a data analyst for the Department for the Economy in Northern Ireland, expresses a similar view. In their view, research data sharing bears great potential in 're-examining' government policies and data custodians should not be afraid of it.[734]

Intermediary bodies, mainly the ones funding administrative data research, have similar incentives to consolidate a perception of administrative data as a shared resource. An interviewee, working for the Economic and Social Research Council in a policy-making capacity, stresses that the promise of administrative data to improve collective knowledge about big, society-level challenges fuelled research investments like the Administrative Data Research Network:

> *'The idea was that we need to exploit the wealth of the UK's administrative data for the purpose of scientific research, in social science and economics*

[731] Interview with participant nr. 33 (2020, virtual).

[732] *Ibid*.

[733] Interview with participant nr. 30 (2020, virtual).

[734] Interview with participant nr. 16 (2018, phone).

(…) enable the linkage of data, admin data, maybe longitudinal data, in order to address the very big societal challenges.'[735]

Interviewees working for intermediary bodies elaborate on the critical role of such bodies in shifting the 'culture' of data custodians towards maximising the use of administrative data for research in the public interest.[736] An interviewee, working in a strategic capacity for Administrative Data Research Centre-Scotland and having experience in successfully negotiating access to census data held by the Scottish government,[737] claims that they have been trying to convince data custodians that 'academics are the conscience of society'.[738] Drawing inspiration from a relevant mention in New Zealand's legislation on academic freedom,[739] they sought to convey the message that evidence-based, constructive criticism of government policies serves the common good. Are, however, these ideas well-received by most data custodians?

While some positive examples have been mentioned above, it is fair to say that, from a data custodian perspective, viewing administrative data as a public good is a work-in-progress rather than an established state of affairs. A researcher affiliated with the Administrative Data Research Centre-Northern Ireland concedes that it is an incremental departure from an understanding of administrative data as being owned by the data custodians. They stress that recent years have seen a 'shift in culture', where administrative data are increasingly seen as 'public data (…) and people want to access it for research purposes'.[740] In not distant times in the past, as an interviewee working as a data analyst for the Knowledge and Analytical Services of the Welsh Government (custodian body) recounts, it would be much more difficult to convince their colleagues to further invest in research data infrastructures like the SAIL databank.[741] They associate this emerging shift in culture not only with day-to-day bargaining

[735] Interview with participant nr. 19 (2018, Swindon).

[736] Interview with participant nr. 10 (2018, Belfast).

[737] National Records of Scotland, 'Welcome to Scotland's Census' <https://www.scotlandscensus.gov.uk/>.

[738] Interview with participant nr. 34 (2020, virtual).

[739] Section 162 (4) (a) (v) of the Education Amendment Act, New Zealand: universities are to accept 'the role of critic and conscience of society'.

[740] Interview with participant nr. 10 (2018, Belfast).

[741] Interview with participant nr. 25 (2019, Cardiff); see Ford et al (n 221) on the SAIL databank and its role in the Welsh research data sharing context.

between researchers and data custodians, but also with a broader increase in appreciating 'evidence-based policy making'.[742]

Other interviewees corroborate the assertion that significant progress has been registered in viewing administrative data as a shared resource and enabling research in the public interest compared to no more than a decade ago. An interviewee, with significant background as both an analyst and a policy-maker for the Office for National Statistics in England, notes this development with appreciation:

> *'For a lot of the UK now, the idea that we should be default open is taken for granted, in a lots of places. Certainly, compared to what we started with many years ago, it's an awful lot easier to get access to data now.'*[743]

In a similar vein, a policy-maker working for the Health and Social Care Department in Wales (custodian body) notes with approval that in recent years people within their organisation have been appreciating the value of administrative data for public service delivery as a whole.[744] Another interviewee working for the Honest Broker Service in Northern Ireland (custodian body) in an information governance capacity explains that their organisation charges researchers for access to data solely because there is no funding available to 'provide the service free of charge',[745] not as a *quid pro quo*. An interviewee, working as a data analyst for the Administrative Data Research Centre-Wales (intermediary body) and having experience in negotiating access to data in Wales, largely concurs. He characterises Welsh data custodians as 'incredibly open' and receptive to criticism:

> *'They just want to know what is the single version of the truth. If it shows things are problematic, they will look to see how they can make it better. And they are interested in what can the data tell them.'*[746]

[742] *Ibid.*

[743] Interview with participant nr. 42 (2020, virtual).

[744] Interview with participant nr. 24 (2019, phone).

[745] Interview with participants nrs. 11, 12 and 13 (2018, Belfast).

[746] Interview with participant nr. 22 (2019, Swansea).

The data presented here from the four countries of the UK demonstrates that perceiving administrative data as a 'public good' and a shared resource facilitates widening access to data for research purposes, even if, *or precisely because*, it allows criticism of government policies. As the next sub-section (ii) will show, there are concrete repercussions for the regulation of research data sharing decisions, i.e. the establishment of transparent procedures and mechanisms of accountability by data custodians who espouse the perception of administrative data as a common good.

ii. Data as the 'public good': establishing mechanisms of accountability

Data custodians who commit themselves to a view of administrative data as a public good, also see themselves as accountable to both the research community and the broader public when it comes to deciding whether they will be granting access to such data. Such data custodians are responsive to feedback from requesting parties and aspire to establish and implement a consistent and transparent data sharing framework. In practical terms, there are certain 'mechanisms' of accountability that underpin such a framework: setting up precedent systems, enabling appeals against a data custodian's decision and delegating decision-making powers to colleagues or trusted third parties with experience in administrative data research. The following discussion will unpack the ways in which perceiving administrative data as a shared resource facilitates the establishment of accountability mechanisms in research data sharing decisions.

Interviewees who are currently working, or have worked in the past, for the Office for National Statistics (ONS) elaborate on the Office's steadfast commitment to becoming an accountable data custodian. Consistency in decision-making is highly valued to ensure that the same standards are applied to different research applications. In the words of an information governance official, it would 'cause us a problem' if similar project requests were treated in a different manner:

> *'Say if we had a project that wants to use labour force survey data (…) and we said yes, and then a year later we had another project request and said no to it,*

> *the second researcher would know that we had approved the other one a year ago because it's published on our website, so what's changed?'*[747]

Another interviewee, working as a data analyst for the ONS, provides some background on the value of consistency in research data sharing decision-making. They explain how consistency almost emerged as a natural requirement within the Office's decision-making practice. As they recall, long before statutory legislation enshrined explicit legal powers for the ONS to share its data for research purposes,[748] the Office felt compelled to establish the Microdata Release Panel (MRP).[749] Until the establishment of the MRP, they admit that decisions were a '*little bit ad hoc'*, relying '*on the memory of the people in the team'*.[750] They identify consistency as the central driver behind 'centralising and formalising' the decision-making process through the establishment of a 'panel of people from around the organisation' (i.e. the MRP).[751] To properly observe consistency and streamline decision-making, the ONS, after careful considerations and lengthy internal deliberations, started developing a system of precedent decisions:

> *'So, over the years, we built up a system of precedent decisions, where we would probably end up about 10% of the requests that were coming in to us actually being considered by the Panel (...) now we have a written set of precedents, which allow us to make decisions and the understanding is that we can make it as a team, make decisions based on those precedents that the panel had agreed.'*[752]

As the interviewee explains here, there is a deciding Panel, the MRP, and a support team, which extracts principles out of the Panel's previous decisions and classifies incoming data access requests according to their similarity with previous decisions (the 'precedents'). If no clear similarity can be established, then the MRP will have to consider the request in full detail. There is a two-fold function of this precedent system. First, internally, it allows better co-

[747] Interview with participant nr. 1 (2017, Titchfield).

[748] The Statistics and Registration Services Act 2007.

[749] See *supra* chapter 4 p 100 on the Office for National Statistics Micro-data Release Panel.

[750] Interview with participant nr. 2 (2017, Titchfield).

[751] *Ibid.*

[752] *Ibid.*

ordination and application of research data sharing decisions, especially when requests are similar to previous requests to use the Office's data. With ca. 250 projects currently facilitated within the ONS Virtual Microdata Laboratory,[753] the precedent system facilitates a more expedient approval process:

> *'We have a lot of these projects, the first one of them gets approved in the full proper process: talking to data owners, going through the panel. Then (…) when the next project comes up, my team says we can use these precedents and I then say yes (…) So, it streamlines the process.'*[754]

Second, externally, it is a practical illustration of the Office's commitment to transparent decision-making, enabling requesting parties to assess whether the ONS's research data sharing criteria are consistently applied in practice. This is not to say that transparency is tantamount to relaxing data access requirements. The Office is always serious about the conformity of proposed research data sharing with strict data security safeguards, including the vetting and training of researchers aspiring to use its data. As an information governance official working for the Office explains, however, the clarity provided to interested researchers by the precedent systems allows researchers to 'try to find another way' of realising their projects even if they are facing challenges in meeting the Office's requirements.[755] Transparency and clarity also allows for better negotiations between the ONS and researchers, who are better-placed to understand the purpose behind the Office's requirements and potentially even challenge them when it comes to e.g. methodological preferences or limitations to the allowed data uses.[756]

Another interviewee, working for the Office as a data analyst, concurs that it is of utmost importance that these criteria are 'transparent and people know exactly what it is that we are doing'.[757] They associate this with a central government push for the civil service to be 'more open and transparent' by widening access to administrative data.[758] In that regard, they consider

[753] Interview with participant nr. 1 (2017, Titchfield); see *supra* chapter 4 p 100 on the development of the Virtual Microdata Laboratory.

[754] Interview with participant nr. 1 (2017, Titchfield).

[755] *Ibid.*

[756] Interview with participant nr. 42 (2020, virtual).

[757] Interview with participant nr. 3 (2017, Titchfield).

[758] *Ibid.*

publishing a so-called 'data prospectus' or 'data catalogue', i.e. a list of available data for research use, as an obligation of all government departments, unless there are specific reasons not to make such a list available.[759]

Apart from establishing a precedent system, the Office's commitment to accountability has been realised through its receptiveness to feedback from the research community, as well as through granting the right to appeal a data custodian decision to requesting researchers. As a policy-maker working for the ONS notes, the Office takes the process of consultation with its data users seriously, without hesitating to amend its data sharing criteria to make them 'clearer and more specific',[760] even if this reduces the interpretative latitude afforded to the decision makers. For example, the Office aspires to follow researchers' feedback and make its current appeals process more formalised. Currently, the ONS adopts an informal approach to allowing appeals against its custodians' decisions. A data analyst working for the Office explains that both their own recommendations and the formal decisions of the Microdata Release Panel can be challenged by requesting parties:

> *'If it was, let's say, my decision, the obvious thing is that the Panel can then reassess my decision. If it is a decision the Panel had already made (...) they would probably wish to reconsider that if there was more evidence that could be provided (...) or maybe a clarification on sth (...) I'm sure that would come back to the Panel and the Panel would just reconsider it.* '[761]

Essentially, this is less of a hostile appeals process and more of a collaborative, constructive effort to exchange views between the researcher and the decision-maker to improve a project request and increase its chances of approval.[762] As the aspiration for widening access to administrative data for research purposes grows, the Office intends to formalise this appeals process to guarantee that requesting parties will be able to effectively challenge custodians' decisions. As the information governance official mentions, the Office aspires to have a sophisticated, two-tier system of appeals in place, with researchers being able to first challenge

[759] *Ibid.*

[760] Interview with participant nr. 4 (2017, Titchfield).

[761] Interview with participant nr. 2 (2017, Titchfield).

[762] Interview with participant nr. 1 (2017, Titchfield).

a decision before the ONS's data governance committee and, ultimately, before the National Statistician (who is the head of the ONS).[763] They even consider the possibility of judicial review after that stage, although they admit that it would be really rare for a project request to reach that stage.

Finally, yet importantly, a critical indication of the ONS's commitment to accountability is the readiness of its data custodians to delegate decision-making responsibilities to colleagues who are closer to using the data for research purposes as data analysts. In other words, policy-makers, who are the ones formally responsible for approval decisions, may delegate such decisions to the technical experts, the data analysts, in spite of the latter being lower in the organisational hierarchy. Delegation, of course, did not happen overnight, but it took a lot of bargaining to be established as a practice. An interviewee who has worked both as a data analyst and a policy-maker for the Office vividly recounts their experience in negotiating such a delegation of decision-making capacity. They recall the example of a particular information asset owner and their efforts in allowing the Office's analysts to perform critical data security assessments with regard to research data sharing:

> *'Because we know what the research is, what we want to get out of this, we said to the (redacted) team, why don't we do this disclosure control? We could make these decisions quickly - we'll send you what we've done, our decisions, and you can check it. You can keep an eye on this, but, basically, we will do the work (...) and the responsibility of signing-off something which you don't really necessarily understand.'*[764]

As the same interviewee recalls, the data custodians agreed to this arrangement and, in fact, exercised a light-touch review over the disclosure control reports submitted to them by the Office's analysts.[765] Although this may look like a minor delegation of technical responsibility, the interviewee finds it significant from a symbolic perspective: it brought a 'change in culture' by gaining acceptance of the position that 'the people who are responsible for the research are

[763] *Ibid.*

[764] Interview with participant nr. 42 (2020, virtual).

[765] *Ibid.*

the best-placed to sign-off the outputs'.[766] If, however, the ONS is the luminary example of establishing accountability mechanisms in the name of maximising the benefits of administrative data research for the common good, how common are such mechanisms in the broader UK public sector?

While the label of 'work-in-progress' is, as mentioned in the previous sub-section, fair, there is an increasing tendency of other data custodians to embrace the values of consistency and transparency in research data sharing decisions, resulting in the establishment of accountability mechanisms. An interviewee who has worked in a policy-making capacity for the HM Revenue and Customs (custodian body) highlights consistency as the driving force behind their department's decision to establish a uniform Data Sharing Strategy[767] and a so-called 'front-door' service:

> *'With a single front door, you could have a consistent treatment, it assures that the governance is correct and (…) you can potentially design the same solution or recycle the same solution to meet multiple needs.'*

Another interviewee, who has worked as a data analyst for the Department for Education (DfE) in England, hails the achievement of the Welsh government to be consistent in their data sharing standards, as opposed to England where 'you have different departments, everyone has their own separate arrangements and legislation'.[768] They express their hope that England will manage to achieve a similar level of consistency in the future, drawing inspiration from the DfE itself. The DfE takes great pride in being consistent and transparent in its research data sharing decision-making and it is, arguably, the ministerial department which publishes the most information about its external data shares.[769] This is admitted by an interviewee who works for an independent research organisation and has worked in the past for the DfE as a data analyst. In their view, the data like the National Pupil Database and the information about

[766] *Ibid.*

[767] HM Revenue and Customs (n 156).

[768] Interview with participant nr. 30 (2020, virtual).

[769] Department for Education (n 61).

the data provided by the DfE to interested researchers is a 'resource like no other in the public sector'.[770]

Beyond being consistent and transparent about their research data sharing decisions, the DfE has been willing to delegate its research data access arrangements to other organisations in recent years.[771] In other words, the DfE have authorised another public body to decide on their behalf about whether particular research data sharing requests should be granted or not. More specifically, they have delegated the responsibility to handle access requests to the Secure Research Service, owned by the ONS.[772] According to one of the interviewee who has worked as a data analyst for the DfE, many other government departments are considering to follow DfE's footsteps and delegate decision-making responsibilities to trusted third parties.[773] The interviewee with the previous experience at HM Revenue and Customs (custodian body) concurs, expressing the aspiration to move towards a model where departments do not 'own' and hold the data themselves, but are entitled to access individual records based on their policy requirements.[774] The image of delegation of custodian responsibilities is corroborated by an administrative data researcher with experience in negotiating data access in Scotland. As they mention, the Scottish government has been aiming to centralise data by handling access to datasets it curates on behalf of its various departments.[775]

Interviewees from the devolved UK countries elaborate on how the mechanism of delegating the responsibility to make data access decisions has been always prevalent in respect of access to health data, which is paradigmatically considered as a common good.[776] An interviewee working in an information governance capacity for the NHS Wales Informatics Service (NWIS) (intermediary body) explains that they have been handling access to research requests 'on behalf and under the instruction of the data custodians'.[777] NWIS, by virtue of their expertise in health data management, are making access decisions on behalf of such data

[770] Interview with participant nr. 33 (2020, virtual).

[771] Interview with participant nr. 30 (2020, virtual).

[772] *Ibid.*

[773] Interview with participant nr. 30 (2020, virtual).

[774] Interview with participant nr. 32 (2020, Oxford).

[775] Interview with participant nr. 35 (2020, virtual).

[776] Understanding Patient Data (n 134); Jones et al (n 214).

[777] Interview with participant nr. 23 (2019, Cardiff).

custodians as National Health Service Health Boards in Wales. When asked whether a research data sharing request would necessitate going back to the data custodians and asking for their approval, they respond in the negative:

> *'No, not necessarily. We are given that element of control in some respects (…) For instance, for any research request, we would look at that request, see if it has been through the normal ethics approval and (…) what are the legal aspects that ensure appropriateness of the data sharing.'*[778]

An interviewee working as a policy-maker for the Health and Social Care Department in Wales (custodian body) adds that the existence of 'secure locations' like the SAIL Databank are integral to convince organisations to sign-off and delegate their custodian duties to decide about access.[779] In the Welsh context, an interviewee working as a data analyst for the Knowledge and Analytical Services (KAS) of the Welsh government notes that such delegation has expanded from health data sharing to statistical research, with the statisticians of the KAS convincing their equivalents within data custodians to 'delegate the permission to share data for statistical research purposes'.[780]

Another interviewee, working as an information governance official for the Electronic Data and Research Information Service in Scotland, clarifies that it is the Scottish government's aspiration to adopt a similar rationale (i.e. delegating access decisions) in the interest of streamlining access to administrative data in the broader public sector. They mention Scottish education data as one of the first examples where this has been made possible, advocating for a 'graded approach in terms of the governance' when it comes to data custodians that may be more reticent to delegate responsibilities.[781] An interviewee, working in an information governance capacity for the Honest Broker Service (HBS) in Northern Ireland, notices a similar trend of delegating custodian responsibilities to other trusted organisations. They point out that they are currently exploring their options with a view to expanding the delegation of access

[778] *Ibid.*

[779] Interview with participant nr. 24 (2019, phone).

[780] Interview with participant nr. 26 (2019, Cardiff).

[781] Interview with participant nr. 36 (2020, virtual).

decisions, from health data sharing where it already happens in Northern Ireland as in Wales (with NWIS), to broader administrative data research.[782]

Beyond the example mentioned here, information is scarcer about the employment of similar accountability mechanisms within the public sector, although it seems that several public bodies are likely to rely on precedent systems and allow appeals of their decisions, even if not formally. An interviewee working as a policy-maker for the Economic and Social Research Council (intermediary body) points out that many data custodians often rely on their previous decisions, in the interest of 'simplifying the complex requirements of the legislative landscape'.[783] A data analyst working for the KAS of the Welsh government suggests that they aim to organise their previous decisions so that they can 'build on them' when a new request is submitted.[784] Another interviewee, working in a policy-making capacity for the HBS in Northern Ireland, explains that their organisation established an appeals route for researchers following the Welsh Health Department's 'commitment to transparency in 2012-13'.[785]

The data presented here illustrate a strong association between viewing administrative data as the 'public good' and establishing accountability mechanisms to ensure that research data sharing decisions will be consistent and transparent. Bodies like the Office for National Statistics in England, as well as particular data custodians in the devolved countries, have been mentioned here as examples of this type of, facilitative for research, data sharing culture. Nonetheless, as it has been shown, this perception of administrative data is by no means unrivalled. As the following sub-sections (iii & iv) will demonstrate, a conflicting perception of administrative data ownership is also salient in influencing the regulation of research data sharing decisions.

iii. Data as 'custodian property': 'our' data and 'their' research

Not all data custodians share a perception of administrative data as a shared, common resource that should be utilised for the broader public interest. To many, administrative data is 'owned'

[782] Interview with participant nr. 12 (2018, Belfast).

[783] Interview with participant nr. 19 (2018, Swindon).

[784] Interview with participant nr. 26 (2019, Cardiff).

[785] Interview with participant nr. 12 (2018, Belfast).

by the custodian body responsible for its collection and management. Such custodian bodies often express strong ownership attitudes, treating administrative data researchers as external parties who are interested in benefiting from their property. As it will be shown, strong perceptions of administrative data ownership are contingent on context and organisational mission, with large ministerial departments like the Department for Work and Pensions (DWP) or the Department for Education (DfE) being paradigmatic examples.

An interviewee who has worked for the DfE in England as a data analyst provides some background to perceptions of administrative data as the property of the department. To them, the nature of the data is important; education data may be seen more naturally as public record compared to data related to income or taxes:

> *'In education, people's data are almost public record, whereas e.g. DWP or HMRC data, it's a far more sensitive thing for a department to be releasing both people's earnings data (…) People expect that if they are reporting their incomes to HMRC, then the HMRC would hold on to that data and not go anywhere else.'*[786]

There is a number of ways in which data custodians express their attitudes of ownership over administrative data, either implicitly or explicitly. One common strategy of data custodians, with a view to ensure that they fully determine the uses of data by researchers, is to reserve the right to unilaterally revoke their consent to the data sharing agreement and, thus, block data access at any time. In that sense, custodians refuse to engage in a bargaining process with researchers as to the scope of data sharing, exercising their prerogative to widen or narrow that scope at different times. An interviewee, working for the Administrative Data Research Centre in Northern Ireland (intermediary body) as a data analyst, recalls that already agreed upon projects often see delays due to last-minute objections by data custodians. Sometimes, these objections are whimsical rather than comprehensively explained, reflecting a strong attitude of entitlement on behalf of the custodians:

> *'In this project, we still don't have a data sharing agreement signed, because there was a change of staff (…) and the new guy didn't agree with the older*

[786] Interview with participant nr. 33 (2020, virtual).

> *guy's policy and was much more conservative and saying ''you are not having it''. And he is still saying ''you are not having it''*[787]

The same interviewee attributes this attitude to the tendency to see administrative data as 'the property of the Department', noting that it is the exception to find data custodians who 'really like research and think it's worth doing'.[788] An interviewee, who has worked as both a data analyst and a policy-maker for the Office for National Statistics, recalls a similar incident which was arguably not representative of the Office's overall facilitative approach to research data sharing:

> *'There was one particular dataset, the person managing it, didn't really like the research side. They were very happy to supply data to the other government departments, but they didn't see the academic community as their customer base and they really resented that.'*[789]

Another interviewee, working as a data analyst for the Department for Education in Northern Ireland, concedes that a data sharing agreement 'does not necessarily guarantee access to data'.[790] Data custodians may revoke their consent to a research data sharing process, without necessarily providing a justification for their withdrawal. This is corroborated by an interviewee, working in an information governance capacity for the Northern Irish Statistics and Research Agency (NISRA). They clarify that a data sharing agreement does not preclude custodians from 'vetoing' projects or particular variables that researchers were interested in accessing.[791] The picture is not different in other UK countries: the DfE in England still reserves the right to refuse access even after having outsourced their data access arrangements to the ONS,[792] Welsh data custodians retain the right to 'review any proposal to use the data'[793]

[787] Interview with participant nr. 15 (2018, virtual).

[788] *Ibid.*

[789] Interview with participant nr. 42 (2020, virtual).

[790] Interview with participant nr. 16 (2018, phone).

[791] Interview with participant nr. 9 (2018, Belfast).

[792] Interview with participant nr. 30 (2020, virtual).

[793] Interview with participant nr. 21 (2019, Swansea).

and Scottish data custodians 'can potentially turn around and say ''no, we are not going to provide the data'''.[794]

Apart from unilaterally changing their mind about providing access to administrative data for research purposes, data custodians also express their perception of data as their property when they take offence at not being consulted during every stage of the data sharing process. An interviewee working for the Administrative Data Research Centre-Northern Ireland (intermediary body) as a data analyst clarifies that custodians 'want to know what comments' other data custodians or intermediary bodies have on the research proposal, often creating an 'endless loop of revisions'.[795] Two interviewees working for the Department for Work and Pensions (DWP) (custodian body) clarify that they see it as both their right and their obligation to be involved in and advise about every stage of the data sharing process. The first of them, who works as a data analyst, claims that the Department 'owes' to people that it will be ensuring that their data is used for a 'worthwhile purpose'.[796] The second, who works in an information governance capacity, agrees that the Department should have the 'final word' at every step of the research data sharing process; relying on external bodies is impossible, unless the DWP can be certain about 'the data processing model that has been approved elsewhere'.[797]

Data custodians express their perceptions of data as their property in their multiple interactions with intermediary bodies and researchers. Interviewees who represent the interests of funders reflect on their own experiences in that regard. An interviewee working for the Welcome Trust (WT) (intermediary body) as a policy-maker expresses some scepticism about the progress towards viewing administrative data as a 'public good':

> *'There is definitely further to go before reaching the point where data will be viewed as a public good and shared resource. There is definitely a dominant view that data is something on which departments have rights on to allow them get credit.'*[798]

[794] Interview with participant nr. 36 (2020, virtual).

[795] Interview with participant nr. 15 (2018, virtual).

[796] Interview with participant nr. 7 (2017, Sheffield).

[797] Interview with participant nr. 8 (2017, Sheffield).

[798] Interview with participant nr. 17 (2018, London).

This is not to say that the WT endorses this view, but the interviewee's words certainly convey an overall understanding of the view as dominant in the research data sharing space. An interviewee working for the Economic and Social Research Council (ESRC) (intermediary body) in a similar, policy-making capacity agrees unenthusiastically. They explain that the Administrative Data Research Network example illustrates that a 'democratic, equal partnership system' between data custodians and researchers has 'lots of problems' as long as data ownership is narrowly interpreted by custodians.[799] A colleague of theirs, i.e. another ESRC policy-maker, points out that custodian requirements from researchers are often arbitrary and inconsistent from request to request.[800] An interviewee working for the Administrative Data Research Centre – Scotland in a policy-making capacity describes the way in which data custodians often exercise their prerogative:

> *'It's very hard to work in partnership with the Scottish government. They are used to being in charge and, sometimes, it doesn't really work that well.'*[801]

Viewing administrative data as their own property, data custodians are more likely to feel entitled to provide less justifications when deciding whether to grant access or not for research purposes. What practical implications does this have for the regulation of research data sharing decisions? As the following sub-section will demonstrate, the particular perception of administrative data as 'custodian property' is associated with organisational resistance to accountability mechanisms such as the ones established by data custodians who view administrative data as a 'shared resource' and a common good.

iv. Data as 'custodian property': resisting accountability

Custodians who feel strongly entitled to decide how 'their' data will be used by researchers tend to prefer informal data sharing arrangements over formalised procedures, avoiding to bind themselves to publicly available decision-making steps. Informality, as it has been recounted elsewhere, encourages researchers to bargain with data custodians to achieve access to data.

[799] Interview with participant nr. 19 (2018, Swindon).

[800] Interview with participant nr. 20 (2018, phone).

[801] Interview with participant nr. 31 (2020, virtual).

During such bargaining processes, however, the power often lies with the one holding the resource under negotiation, i.e. the data.

One of the most common indicators of data custodian entitlement is when custodians are not inclined to grant the right to appeal an access decision to requesting parties. A policy-maker working for the Department for Work and Pensions (DWP) (custodian body) concedes that there is no appeals process available to researchers:

> *'Nothing in place at the moment (…) there's not really anyone to which you could really appeal and say could you just have a look what they've decided.'*[802]

They propose that researchers, when receiving a rejection, take the opportunity to informally discuss with the decision-makers the aspects of their proposal that were deemed problematic, but clarify that this shall not be seen as a 'formal appeals or reconsiderations structure'.[803] An information governance colleague of theirs at DWP agrees that there is scope for discussion and constructive responses to custodian decisions, yet, in case the decision-makers reconsider, it shall be seen as 'additional information' that allowed a decision to be reached, rather than an amendment to a previous decision.[804]

To what extent does this reluctance to provide an appeals process reflect broader practices in the UK public sector? This is the case more often than not, according to an interviewee who works for the HM Revenue and Customs (custodian body) in a policy-making capacity. They associate this reluctance with ministerial departments, which often require the ministers to approve before granting access to their data. In such cases, allowing an appeal against a ministerial decision would provide scope for proving that government ministers judged erroneously a particular request. This turns the question of appeals into a political issue, which 'makes it a bit more complicated and whimsical'.[805] An interviewee who works in an information governance capacity for the Northern Irish Statistics Research Agency (custodian

[802] Interview with participant nr. 8 (2017, Sheffield).

[803] *Ibid.*

[804] Interview with participant nr. 8 (2017, Sheffield).

[805] Interview with participant nr. 32 (2020, virtual).

body) in Northern Ireland notes that 'nobody has tried to formally appeal or complain against an owner decision'.[806] Another interviewee, who works as a data analyst for an independent research organisation and having experience in negotiating access to Department for Education (custodian body) data, concurs. Asked whether they have ever challenged a custodian decision, they admit that:

> *'Mostly you have to conform to what is being asked (...) I don't think we could not find a way to work around some things (in doing the research), there's always been a solution.'*[807]

Administrative data researchers with experience in negotiating access with Scottish data custodians corroborate the widespread lack of appeals processes. Three researchers who are affiliated with the Administrative Data Research Centre – Scotland (intermediary body) concede that:

> *'To challenge a data custodian decision would be very difficult - they are data controllers, they are really all powerful.'*[808]

> *'They [data custodians] are really powerful - we would pretty much agree to whatever they said, because we wanted the data.'*[809]

> *'I don't know how you would ever come back on that, if someone said we aren't going to give you the data (...) You wouldn't really have a choice I guess.'*[810]

Instead of trying to formally appeal a decision, the first of these three researchers suggests that one sees a data custodian decision 'very much like a peer-reviewed paper', seeking to 'respond constructively to suggestions with which we disagreed'.[811] Nonetheless, they do admit that this is not always successful. They describe the process as 'extremely messy' and 'very bumpy',

[806] Interview with participant nr. 9 (2018, Belfast).

[807] Interview with participant nr. 40 (2020, virtual).

[808] Interview with participant nr. 34 (2020, virtual).

[809] Interview with participant nr. 35 (2020, virtual).

[810] Interview with participant nr. 29 (2020, phone).

[811] Interview with participant nr. 34 (2020, virtual).

with researchers feeling that they do not know 'where things are at'.[812] When researchers are left in the dark, it is difficult for them to negotiate and bargain with data custodians about the scope of their access request in an effective manner. In one case, their request to get access to specific variables was rejected without a convincing justification:

> *'There were variables that we would very much like to have used, but one individual was able to stop us from getting it because they thought it would be too difficult for us to analyse and I think that's ridiculous that one person can hold so much power (…) you need one person, it knocks everything you've been building.'*[813]

Refusing researchers the right to appeal a custodian decision is indicative of the entitlement a custodian body feels over their data. It is not, however, the only indication of this entitlement. Before the actual stage of a decision, many custodians express their ownership entitlements by refraining from publishing any or sufficient information about their data sharing procedures, effectively keeping requesting parties and intermediaries in the dark. An interviewee, who currently works for an independent research organisation and has worked for the Department for Education (DfE) (custodian body) as a data analyst in the past, admits that they are able to navigate the Department's data access requirements because of their past experience working for it:

> *'When we started applying, it was a little bit like having to guess (…) The advice from DfE would be quite limited on that and we had to guess (…) The information would not really flow appropriately many times (…) If you do it for the first time, then good luck, I don't think you will necessarily navigate this, at least not after a long time.'*[814]

Another interviewee, who works for an independent research organisation as a data analyst, agrees. They complain that they do not find the DfE's data sharing process 'at all

[812] *Ibid.*

[813] *Ibid.*

[814] Interview with participant nr. 33 (2020, virtual).

predictable'.[815] As they explain, the default position is that the decision-makers do not provide much information about the decision-making process, with the researchers having to 'chase it along'.[816] Even after two decades of experience in negotiating access to DfE data, the interviewee concedes that they still have 'no idea' about the different roles of the officials making data access decisions on the DfE's behalf.[817] This may come as a surprise considering the previous assessment of the DfE as one of the most transparent departments in the research data sharing space,[818] as well as considering that education data may be relatively less sensitive compared to e.g. income data.[819] In fact, the uncertainty expressed by the interviewee is indicative of the broader confusion about the roles of different officials (data analysts, information governance, policy-makers) within decision-making.

In a similar vein, an administrative data researcher, who is affiliated with the Administrative Data Research Centre-Scotland, reflects on their experience in negotiating access to Department for Work and Pensions (DWP) (custodian body) data. They recall that, during negotiations, they were 'not so clear' on the Department's requirements and the progress of the request, even though they had regular meetings with DWP officials. Interestingly, the interviewee does not attribute responsibility for the lack of transparency to the DWP analysts who were present in the meetings: these analysts were found to be 'very helpful and progressive'.[820] Delays seemingly 'had to do with other dependencies (…) it was difficult to gauge what the priorities were'.[821] This is consistent with the experience of another interviewee, who has negotiated access to DWP data on behalf of the Administrative Data Research Network (intermediary body):

> *'If I would have to describe my relationship with them: it was always very helpful and very pleasant with the people, but always very frustrating with the organisation.'*[822]

[815] Interview with participant nr. 40 (2020, virtual).

[816] *Ibid.*

[817] *Ibid.*

[818] *Supra* (n 61).

[819] Interview with participant nr. 33 (2020, virtual).

[820] Interview with participant nr. 38 (2020, virtual).

[821] *Ibid.*

[822] Interview with participant nr. 5 (2017, Essex).

The problem, from the interviewee's perspective, was that analysts were only 'disclosing information partially' about the request's progress,[823] making it challenging for them to estimate why things are not moving forward:

> *'I still haven't got the full picture of how many (decision-making) layers are there, because (…) there was only partial information that was disclosed (…) When we were lifting the first barrier, we didn't know how many barriers we would have until we reached that point.'*[824]

DWP decisions, in this interviewee's narration, resemble a Russian doll: the more barriers researchers were overcoming, the more they were faced with novel challenges and unknown difficulties. The same interviewee clarifies that this is far from exclusive in the case of the DWP. In their experience of facilitating Administrative Data Research Network projects in the UK, there were many cases were there was 'no information from the data *owners*, no interaction' (emphasis added).[825] In many cases, custodians do not feel obliged to provide reasons for delaying their response to a research request, nor do they necessarily provide a timeframe within which the request will have been dealt with. These delays, however, can be seriously detrimental to the realisation of an academic research project. As an administrative data researcher who is affiliated with the Administrative Data Research Centre in Northern Ireland (ADRC-NI) eloquently puts it:

> *'The frustrating thing for us is (…) that everybody sitting on the academic side of the wall doesn't exist - working in a project, getting a project funded, getting it written, published - we had researchers which (…) needed to produce outputs.'*[826]

[823] *Ibid.*

[824] *Ibid.*

[825] *Ibid;* even the word 'data owners' used by the interviewee is indicative of the attitudes of entitlement expressed by data custodians in this relationship.

[826] Interview with participant nr. 10 (2018, Belfast).

Colleagues of theirs from the ADRC-NI agree that there is often a lack of 'forward facing information to researchers',[827] making it difficult for researchers to plan their studies. Another researcher, with experience in negotiating data access with Scottish data custodians, recounts their own difficulties in understanding what is required from them:

> *'We kind of assumed that once everybody said yes and signed some contracts, that was going to be it. We did not realise it would be a whole process of actually negotiating the contracts, negotiating the particular conditions of access etc. There was a whole kind of middle section that we were not aware of.'*[828]

They complain that most (non-health) data providers in Scotland do not have 'any discoverability for their datasets',[829] making it essential that interested parties engage in a relationship with specific persons within the data custodian in the interest of understanding the requirements and, ultimately, getting access. This relationship often requires delicate balancing between the priorities of researchers *vis a vis* the priorities of custodians, an exercise that is not always facilitated by the information provided to researchers by custodians. An interviewee working for the Economic and Social Research Council (ESRC) (intermediary body) in a policy-making capacity confirms that the most common complaint they get from their funded researchers is that the latter 'never get a clear answer back' from data custodians with regard to the status of their access request.[830] Another interviewee, working in a similar capacity for the Administrative Data Research Council (ADRC) (intermediary body) in Scotland, admits that researchers are 'usually astonished at how long it takes and how much they actually have to do'.[831]

It is not only researchers that are kept in the dark by custodians. The latter's attitude of entitlement over the data is also expressed in their interactions with intermediary bodies. The same interviewee from the ESRC recalls that, during the early stages of the Administrative Data Research Network (ADRN) (intermediary body) investment, they particularly struggled

[827] Interview with participant nr. 15 (2018, virtual).

[828] Interview with participant nr. 29 (2020, phone).

[829] *Ibid.*

[830] Interview with participant nr. 19 (2018, Swindon).

[831] Interview with participant nr. 31 (2020, virtual).

with convincing custodians to provide them with some information on why project requests were being blocked.[832] ADRN members often had to engage in long-drawn conversations with custodian members without satisfactory outcomes. An interviewee, who has negotiated access to Department for Work and Pensions (DWP) (custodian body) data on behalf of the ADRN, explains that it is not easy to press data custodians to disclose more information about their decision-making process:

> *'If you start to get stuck on something, and then you start to ask questions like, so who is your boss exactly? You know, how do you do that? That's a difficult kind of conversation to have with someone.'*[833]

Considering the 'perceptions of administrative data ownership' on the whole, the data from the four countries of the UK presented in this section illustrate the association between particular attitudes and the establishment of accountability mechanisms and procedures, or lack thereof. Custodians who are committed to viewing administrative data as a shared resource are likely to introduce transparent data sharing procedures and accountability mechanisms. They will do so in the interest of achieving transparency and consistency in the way in which they manage access to a public good. By contrast, custodians who feel that administrative data belongs to their organisations are much more likely to express their strong feelings of entitlement by refusing to provide reasons behind their decisions, as well as relying on more informal and less accountable data sharing processes. While perceptions of data ownership are very strong in influencing the formation of data sharing frameworks in the UK public sector, there is another significant constituent of the dominant research data sharing culture: policy relevance of administrative data research.

B. Policy Relevance of Administrative Data Research: 'What is in it for us'?

Like perceptions of administrative data ownership, assumptions about the policy relevance of administrative data research are unwritten, uncodified and non-articulated. Yet, as it will be discussed, they are very influential in the process of establishing and implementing research data sharing frameworks in respect of most data custodians. Although there are some notable

[832] Interview with participant nr. 19 (2018, Swindon).

[833] Interview with participant nr. 6 (2018, phone).

exceptions like the Office for National Statistics (custodian body) in England, whose organisational mission is to promote data-driven research and knowledge, most data custodians evaluate the alignment of a proposed research request with their policy priorities and their capacity to deliver them.

How does this shape the 'rules of the game' that filter access to administrative data? When custodians are satisfied that research satisfies their policy aims, they are more likely to set up a data sharing process, even committing themselves to specific decision-making steps. When, however, they are unclear as to the relevance or the compatibility of a proposed research request with their policy agendas, inertia, reluctance and caution are likely to emerge, hampering progress in the establishment or the transparent implementation of data sharing frameworks. This section proceeds from providing some background on the construction of this constituent of the research data sharing culture, before highlighting the ways in which it concretely accelerates or hampers the establishment of data sharing procedures.

i. 'Blue skies' research and the normative supremacy of policy-relevant research

As discussed when presenting the historical development of the research data sharing space in the UK,[834] the potential of data-intensive research to make government policies more 'efficient' was a significant driving force behind widening access to such data for external researchers. Administrative data research allows public bodies to 'measure the effects of policies and spend government money in a more effective way'.[835] This allure of policy-relevant data-intensive research has seemingly consolidated a conventional hierarchy between 'applied' and 'theoretical' research in the mindset of many data sharing participants. The former has come to be regarded as more appropriate and significant than the latter.

An interviewee, who has worked for the Department for Education as a data analyst, explains that, from a custodian perspective, the most significant reason for approving research data access requests is the promise of 'enhancing the evidence-base, meaning that we can

[834] *Supra* chapter 4, B.

[835] Interview with participant nr. 3 (2017, Titchfield).

improve policies'.[836] Interviewees working for the Department for Work and Pensions (DWP) strongly share this view. One of them, who is working for the Department in an information governance capacity, explicitly dismisses research that cannot help evaluating DWP policies. They go as far as implying that non-relevant research questions cannot qualify as research in the public interest:

> *'We can sometimes see requests for data sharing for something that demonstrates a very interesting methodological question, but of no practical use whatsoever and we have to legitimately ask the question: is that in the public interest? What public interest would then be beyond any public sector body? (…) is this in the public interest in terms of the benefit of the country or a ''nice-to-know'? '*[837]

Essentially, the interviewee here defines 'public interest' research by reference to the policy priorities of specific organisations, either the DWP or other departments. They explain that they find this a justified position because the taxpayers would expect the DWP to invest their resources into research that is directly relevant to its functions.[838] A colleague of theirs, working as a policy-maker for the DWP, confirms that a request's contribution to an 'evidence gap' or to something that the Department would 'positively want to see done' is crucial.[839] They calmly add that 'there is no way' that the Department would provide sensitive information to a researcher 'just cause they think it will be interesting'.[840] Unless the policy relevance of the request can be demonstrated, they add, the 'value of research is not quite as obvious' to the department's decision-makers.[841] An administrative data researcher, who has negotiated access to DWP data, describes 'an alignment of interests' between the researchers and the Department

[836] Interview with participant nr. 30 (2020, virtual); it is important not to conflate 'evidence-based' with 'policy-relevant', even if the two terms are often used together. The Office for National Statistics is keen to promote evidence-based research, but as it has been shown, is less strict on the policy relevance compared to departments like the Department for Education or the Department for Work and Pensions.

[837] Interview with participant nr. 8 (2017, Sheffield).

[838] *Ibid.*

[839] Interview with participant nr. 7 (2017, Sheffield).

[840] *Ibid.*

[841] *Ibid.*

as crucial in moving things forward.[842] A data analyst, who has negotiated access with the DWP on behalf of the Administrative Data Research Network, summarises the Department's approach as follows:

> *'if a project doesn't fit with what we're interested in and researching these priorities this time, then we would probably not devote resources to it.'*[843]

As is often the case in the research data sharing space, the HM Revenue and Customs (HMRC) (custodian body) is not radically different from the DWP. An interviewee who has worked in a policy-making capacity for the HMRC explains that a researcher using the HMRC DataLab 'almost undertakes one of the functions for HMRC'.[844] In other words, researchers aspiring to use HMRC DataLab data will effectively work on research questions that form part of the Department's own policy-making priorities. They mention such areas as compliance or taxation around novel areas of commercial activity as particular examples of research that would be likely seen in a positive light. The same interviewee captures the distinction drawn within their department's data sharing criteria between policy-relevant and non-relevant, academic research:

> *'So, we factored if: it needs to be critical to major transformation for a department; if it was critical to delivering a new policy; critical to a minor academic project; does it improve ease of delivery; shorten the cost or delay to pay a social security benefit etc.'*[845]

Coming to the devolved countries, an information governance official working for the NHS Wales Informatics Service (intermediary body) draws a distinction between 'small, isolated' studies that are supported e.g. by one University and 'much wider sponsored' studies through e.g. the Welsh government.[846] Another interviewee, working for the Department for the Economy in Northern Ireland (custodian body) (DfE-NI) as a data analyst, describes a similar

[842] Interview with participant nr. 38 (2020, virtual).

[843] Interview with participant nr. 5 (2017, Essex).

[844] Interview with participant nr. 32 (2020, Oxford).

[845] *Ibid.*

[846] Interview with participant nr. 23 (2019, Cardiff).

picture. They concede that the 'policy relevance of the work' is one of the primary considerations for departments. As they put it, when considering research data access requests, data custodians often need to see 'what is in it for them':

> *'From first-hand experience of being in a meeting, asking them to provide data for specific purposes, being told that they could not see where there was any value for them. I think it's very much in the front of their mind in terms of ''what's in it for us''.'*[847]

The same interviewee recalls a particular incident, during which the DfE-NI had allowed access to their data on the condition that a group of academics would do some economic analysis related to education and household data.[848] The DfE-NI themselves were interested in this analysis for their policy work. When the final report came out, however, there was strong 'dissatisfaction' in DfE-NI custodians because the researchers had not addressed the particular type of question that the Department was interested in, making it of no use to the custodian policy-makers.[849] As this example shows, bargaining to access data may eventually result in conflict between custodians and researchers, when e.g. custodians believe that they are essentially commissioning research but researchers feel that this is their independent research project. Another interviewee, working as a data analyst for the Knowledge and Analytical Services (custodian body) of the Welsh government, remembers how important it was for them to demonstrate the policy relevance of a proposed research project to get their higher-ups to establish a data sharing process:

> *'In the (redacted) case, it was the key reason that I managed to get the funding out of policy - that big evidence gap was being able to evidence health impact. Once they were aware that this was possible, that there was even a thing they could do, they were like ''this is amazing''.'*[850]

[847] Interview with participant nr. 16 (2018, phone).

[848] *Ibid.*

[849] *Ibid.*

[850] Interview with participant nr. 25 (2019, Cardiff).

While this was a case of a researcher's success story, it is often the case that the perception of policy-relevant research as more appropriate complicates data access. A policy-maker, working for the Administrative Data Research Centre (ADRC) (intermediary body) in Scotland, reflects on the difficulties of convincing data custodians about the alignment of independent research projects with their policy aims. They express strong disappointment at the lack of a more open-minded approach on behalf of the data custodians:

> *'It's a fundamental problem across getting data and having the results used by policy-makers. The Government is often only interested in research it has commissioned itself and it can control itself!'*[851]

To the interviewee, administrative data researchers will always struggle to convince custodians because the latter 'want an answer to something that week, or that month',[852] whereas administrative data research is about building a comprehensive and long-term evidence base. This may often result in the researchers seeking to overstate the policy relevance of their proposed researcher. A data analyst who has worked for the Administrative Data Research Network (intermediary body) recalls this as common practice:

> *'What happens is applicants tend to, you know, really make sure they (…) make some strong statements about the benefits of doing this piece of work (…) The reality is the method or the data might not allow that definitive clarity, so you can't make the claim that you'll be able to achieve that. So, you have to kind of scale that back.'*[853]

A data analyst working for the Administrative Data Research Centre (intermediary body) in Northern Ireland reflects on a relevant incident, shedding light on the previously mentioned dissatisfaction of the Northern Irish Department for the Economy (custodian body) with a research project that used education and household data.[854] As the interviewee, who was a member of the requesting researcher group, recalls, the researchers were keen on accessing the

[851] Interview with participant nr. 31 (2020, virtual).

[852] *Ibid*.

[853] Interview with participant nr. 6 (2018, phone).

[854] *Supra* p 68.

data and, thus, agreed on addressing particular, policy-relevant research questions that were 'tricky' to address in the absence of certain variables.[855] Presenting research as necessarily policy-relevant resulted in mutual dissatisfaction at the end of the project, yet allowed the data sharing process to happen due to policy-relevant research's perception as appropriate and desirable.

One might expect that this perception of policy-relevant research as hierarchically superior is only shared by data custodians. After all, it is *their* policies that can be evaluated through administrative data research and it would make sense for them to be primarily interested in this potential. This view, however, is generally, quite widespread in the research data sharing space, even shared by individuals representing intermediary bodies and researchers themselves. An interviewee, who has worked as an administrative data researcher in the past and works now for the Economic and Social Research Council (intermediary body), reflects on their experience. With an approving tone, they recall that all of the projects they had been working for had a 'very strong policy focus; it wasn't blue skies research'.[856] They believe that this was crucial in them managing to agree access with the data custodians. A policy-maker, working for the Administrative Data Research Centre (intermediary body) in Scotland, strongly agrees:

> *'You are never going to do social science sitting in an ivory tower. You are going to have to engage with policy makers.'*[857]

In a similar vein, a policy-maker working for the Administrative Data Research Network (ADRN) (intermediary body), reflected on the Network's difficulties in gaining sufficient data access by noting that its theoretical focus was perceived by some departments as a 'unicorn factory'.[858] Two interviewees, who have worked in a consulting role for the ADRN and the Administrative Data Research Centre in Wales, respectively, attribute the ADRN's shortcomings to what they see as a *naïve* approach as to how important the policy relevance of research is:

[855] Interview with participant nr. 15 (2018, virtual).

[856] Interview with participant nr. 19 (2018, Swindon); the term 'blue skies research' refers to scientific research without immediately apparent 'real-world' applications.

[857] Interview with participant nr. 31 (2020, virtual).

[858] Interview with participant nr. 6 (2018, phone).

> *'This was very much researcher-led, with the idea that, obviously, this will provide benefit to the DWP. But government departments don't work like that, there's a lot more need to recognise how you are getting the buy-in from the users, you need to have a particular customer in mind for this research.'*[859]

> *'It was a learning process, it would literally be ''can we please have data?'' – ''ok, what's in it for us?'' – ''well, nothing, we are academics and we need the data'' - it was quite naive and it developed over time.'*[860]

The second of these two interviewees, who is affiliated with the ADRC in Wales, is categorical in their views that policy-relevance should be a *sine qua non* prerequisite of a data access request:

> *'The old approach of an academic asking for data, there is no benefit for the data owner in an environment where their resources have been squeezed, they are very busy, so why would they make time to give this academic data when they see no benefit?'*[861]

These views indicate not only how widespread this view is, but also how much policy-relevance has become normalised as an appropriate criterion for data access, even from the perspective of administrative data researchers. A policy-maker, working for the Wellcome Trust, explains this by reference to the nature of the policy-making job regardless of whether one represents a custodian body or e.g. an intermediary body: they are constantly 'in a massive rush and doing ten things and want something really clear and oversimplified',[862] as opposed to academic interest in broader questions. A data analyst, working for the Knowledge and Analytical Services (custodian body) of the Welsh government, stresses that it may be the case that academic and policy interests overlap, but that will certainly not be the majority of cases.[863]

[859] Interview with participant nr. 42 (2020, virtual).

[860] Interview with participant nr. 22 (2019, Swansea).

[861] *Ibid*.

[862] Interview with participant nr. 18 (2018, London).

[863] Interview with participant nr. 26 (2019, Cardiff).

Another data analyst, who works for the Administrative Data Research Centre in Wales, has seemingly heeded the call to make research more relevant to policy-makers. They refer to the Administrative Data Research's (intermediary body) mission as 'to work with Government to answer pressing societal questions using data'.[864] They explain that this is a mutually beneficial arrangement for both parties:

> *'As the project progresses, the Government gets emerging findings that can support their policy-making decision and then the academic can also produce an academic report based on that work.'*[865]

Other researchers are also of the same mind. An administrative data researcher affiliated with the Administrative Data Research Centre in Northern Ireland stresses that 'good research should be written in a way that a Department could say, you know, that's a really good question'.[866] Another administrative data researcher, also working for the SAIL Databank in a policy-making capacity, brings up the example of clinical research to make the case for the significance of policy-relevant research. They suggest that the input of the practitioners, in this case clinicians, is important for the research to have any value:

> *'Without the clinical input, it is not clear that the research questions will be right in the first place (…) without reference to the clinician, the findings might be brilliant, but they may not be of value.'*[867]

Another researcher, who has also worked as a data analyst for the ADRN, concedes that aligning research with a data custodian's policy priorities is often a helpful way of satisfying the 'abstract' requirement of serving the public interest:

[864] Interview with participant nr. 22 (2019, Swansea).

[865] *Ibid.*

[866] Interview with participant nr. 10 (2018, Belfast).

[867] Interview with participant nr. 21 (2019, Swansea).

> *'It is so abstract, public benefit. It's like wishing for ''world peace''. It's brilliant and everybody agrees, but it's so abstract and not with tangible financial benefits.'*[868]

They add that through this exercise it is easier to measure how a particular research project can produce public benefits. This is corroborated from the side of data custodians. An interviewee working as a data analyst for the Knowledge and Analytical Services (custodian body) of the Welsh government confirms that it is important for custodians to hear 'an example that's really close to home',[869] if they are to be convinced about the relevance of the request to their policy aims. When that happens, they 'suddenly get their wallets out' and are willing to invest into the data sharing.[870] A similar view is expressed by an administrative data researcher who has negotiated access with Police Scotland:

> *'Turning a general interest in a topic into a specific set of research questions that can be applied to the dataset that people were describing was really difficult, we never resolved that issue.'*[871]

Proceeding from 'what the data provider had actually wanted' was actually easier.[872] A policy-maker, working for the Economic and Social Research Council (intermediary body), agrees that aligning research with specific policy aims is helpful in demonstrating that the research serves the 'public interest' due to the nature of non-health administrative data. With health data, the potential benefits for public health are much clearer e.g. improving healthcare, whereas the benefit to the public through e.g. social or educational policies is more complex due to the latter being 'politically or ideologically loaded in a way that health is less so'.[873]

Not all researchers, however, are equally happy with accepting that policy-relevant research is inherently more appropriate. An administrative data researcher, who is affiliated with the Administrative Data Research Centre (intermediary body) in Scotland, expresses scepticism

[868] Interview with participant nr. 5 (2017, Essex).

[869] Interview with participant nr. 26 (2019, Cardiff).

[870] *Ibid.*

[871] Interview with participant nr. 35 (2020, virtual).

[872] *Ibid.*

[873] Interview with participant nr. 20 (2018, phone).

about the need to demonstrate specific policy impact in order to get access to data. In their view, even if access to data is achieved in the short term, nothing happens to shift the custodians 'attitude to data sharing' in more general terms in a way that will facilitate wider access to administrative data for social researchers in the future.[874] Are these worries justified? To reflect on the impact of the perceptions of policy-relevant research as more appropriate, the following sub-section unpacks their concrete impact on the creation and implementation of data sharing frameworks in the UK public sector.

ii. Policy-relevance within data sharing frameworks: brigading research projects

The unwritten, assumed superiority of policy-relevant research has increasingly been reflected in the creation of data sharing frameworks, shaping the conditions under which data custodians allow access to administrative data. With researchers increasingly grasping the need to incentivise data custodians by demonstrating the policy relevance of their projects, there are two main strategies to facilitate 'appropriate' and 'legitimate' research that have emerged. First, data custodians have been *brigading* research projects, organising requests around certain themes that align well with their policy priorities. *Brigading* is used here to convey the strictness with which the preferrable research themes need to be united and underpinned by the core policy interests of the data custodians. These themes operate as proxies for the notion of the 'public interest' and effectively filter access to administrative data for research purposes. Second, data custodians have been *co-designing* research projects, demanding that they are involved in the construction of the research questions and, in many cases, the research methodology. Alluding to popular catchphrases like 'research co-production' and 'research with impact', also perceived as important for high-quality research outputs, data sharing frameworks, by requiring co-design, seek to steer independent research projects in a direction that is beneficial for custodians.

To start with brigading research projects, data sharing frameworks are structured around a few key areas of policy interest to the particular data custodian. A policy-maker working for the Department for Work and Pensions (DWP) (custodian body) explains that, to their

[874] Interview with participant nr. 29 (2020, phone).

department, it is 'wasteful' to supply data for 'individual, isolated' research projects.[875] If the data sharing policy is focused, instead, on some key themes, then future requests will require 'very similar datasets, a little bit further down the line' and the DWP's analysts will be able to provide them easily by just 'rerunning the code and producing a slightly updated dataset'.[876] An administrative data researcher, who has negotiated access with the DWP, confirms that the department has communicated to them a document with 'areas of research interest',[877] within which they had to frame their research project. Another interviewee, who has worked for the Office for National Statistics (ONS) (custodian body) as a data analyst, explains that not brigading projects around certain key areas means that 'you're going to have a large cost for one project and that project is not clearly going to be of benefit to the organisation that is spending the money on that'.[878]

Where do these 'key areas' come from? An interviewee, who has worked for the ONS as a data analyst, provides some background on the construction of the key themes around which a data sharing framework is developed. They clarify that a data custodian's 'list of strategic objectives' is an essential point of reference for developing such key areas of research interest.[879] These objectives may include boosting productivity and welfare, combatting inequality and unemployment etc. From these broad goals, the custodian's analyst would develop some distinct directions of a research agenda, potentially adding some concrete research questions that are of interest to the department – e.g. inquiring into what is the relationship between unemployment and reoffending.[880] The researchers would, then, have to make a compelling case for how their projects are well-aligned with the data custodian's main areas of interest. Will this be assessed rigorously? According to the interviewee, yes and no. On the one hand, it is important to clearly explain the relevance of research to policy-making. On the other hand, it may be the case that indirect evidence of serving the custodian's policy aims will suffice. In their words:

[875] Interview with participant nr. 7 (2017, Sheffield).

[876] *Ibid.*

[877] Interview with participant nr. 38 (2020, virtual); it is striking that this document was not available in the public domain until 2018.

[878] Interview with participant nr. 42 (2020, virtual).

[879] *Ibid.*

[880] Jones and Elias (n 338).

> *'It's not necessarily that you will directly prove that you will raise productivity, but it's to show awareness that you know this is what they have been trying to achieve (…) I can't tell you that this piece of work has changed your goal, but it's clearly contributing to the strategic goal as a body of work.'*[881]

A policy-maker, who has worked for the HM Revenue and Customs (HRMC) (custodian body), reflects on the creation of their department's 'data sharing strategy' and the need to concentrate focus on specific areas of policy interest. They remember that the production of this strategy was the product of lengthy internal organisational deliberations, which often involved delicate balancing of the interests of the different parts of the organisation. As they recall, the Executive Committee of the HMRC had identified a risk in the department 'not exploiting data as best or as effectively as it could do',[882] due to different lines of duty within the HMRC adopting different priorities when sharing data. The strategy was devised to ensure 'a more strategic approach from a departmental view' with regard to using the HMRC's data assets and produce 'genuine public benefit for the department and other departments'.[883] Compliance with tax regulations, managing debt and counter-fraud were some of the key areas of interest, with researchers having to make the case about how their projects would contribute to the HMRC's understanding of these policy issues.

Data custodians in the three other UK countries seem to be even more familiar than English custodians with the practice of requiring conformity with specific research 'themes' within their data sharing frameworks. An information governance official, working for the Administrative Data Research Centre – Northern Ireland (ADRC-NI) (intermediary body), clarifies that their organisation had reached 'agreements in principle' around specific areas of research interest with Northern Irish data custodians, requiring that researchers who want to access data are 'interested in that kind of data'.[884] Another interviewee, who has worked for the same body in a policy-making capacity, explains that such a thematic approach to research data sharing has been a constant element of the data sharing landscape in Northern Ireland. They explain that there were four main, broadly conceived, areas of interest for the Northern

[881] Interview with participant nr. 42 (2020, virtual).

[882] Interview with participant nr. 32 (2020, Oxford).

[883] *Ibid.*

[884] Interview with participant nr. 9 (2018, Belfast).

Irish government (as a data custodian): 'migration', 'education', 'benefits' and 'health'.[885] As a colleague of theirs, working for the ADRC-NI as a data analyst, explains, research requests would have to fall under these broad areas, even if meaning that 'academics are not going to have as free a hand in determining what kinds of projects must be taken forward'.[886] The idea has traditionally been, as another data analyst working for the ADRC-NI notes, that independent researchers will 'make do' with what is available:

> *'Once the dataset is available, researchers will find something to do with it. They will find a corner that hasn't been looked at. If the datasets are curated and kept, they will do something with it.'*[887]

Like their Northern Irish counterparts, Welsh data sharing participants do seem privy to the significance of brigading research projects around certain key areas of policy interests. An interviewee, who works in a policy-making capacity for the SAIL Databank, provides some background to the historical establishment of the databank back in 2005. They recall that the Welsh government were keen on promoting 'thematic research areas', through 'infrastructure support services' that will be able to stimulate ongoing engagement with these areas.[888] The SAIL Databank was funded by the government as such an infrastructure. As the interviewee explains, a similar 'themed' approach is currently followed by the Administrative Data Research Centre (ADRC) in Wales (intermediary body) around the following areas of policy interest: 'housing', 'homelessness', 'mental health' and 'children'.[889] They explain that the thematic approach allows Welsh data custodians to see that research can produce 'clear benefits' for them, in the sense of a 'return of their investment' in putting in resources to enable research data sharing to happen.[890] A colleague of theirs, working for the ADRC in Wales as a data analyst, confirms that it would be the 'standard research approach' to:

[885] Interview with participant nr. 10 (2018, Belfast).

[886] Interview with participant nr. 14 (2018, Belfast).

[887] Interview with participant nr. 15 (2018, virtual).

[888] Interview with participant nr. 21 (2019, Swansea).

[889] *Ibid.*

[890] *Ibid.*

> *'Undertake a thematic research project, produce a research paper and (…) then take that paper to a policy area and say are you interested.'*[891]

Welsh data custodians, the data analyst explains, see this as 'very much a quid pro quo':

> *'I am approaching you, you are a government department, I want to know what projects would benefit you. I will work with you on those projects, you will share data with me to do those projects and at the end we will establish a relationship where you are happy to provide data in the future.'*[892]

Making certain areas of policy interest public enables this research data sharing collaboration to happen. Interviewees working for the Welsh government provide further justifications for this approach. A policy-maker, working for the Health and Social Care department (custodian body) of the Welsh government, claims that 'themes' provide the necessary accountability to evaluate 'how a good research system needs to run and support academic ventures in the long term rather than on a project-by-project basis'.[893] Themes provide the necessary coherence and overarching logic, by reference to which the contribution of individual research projects to collective knowledge will be assessed. A data analyst, who works for the Knowledge and Analytical Services (custodian body) of the Welsh government, adds that this coherence is integral to make research seem 'serious' and appeal more to all stakeholders, including higher-ups in the government.[894] They note that this ability of administrative datasets to be organised around certain themes and then be curated, retained and reused in future research is their greatest strength compared to other sources of data like surveys. As they explain:

> *'Ministers are more likely to fund something if there is an ongoing admin infrastructure underneath it than just a one-off survey or even repeated surveys after certain intervals. They can see the benefit of having an admin*

[891] Interview with participant nr. 22 (2019, Swansea).

[892] *Ibid*.

[893] Interview with participant nr. 24 (2019, phone).

[894] Interview with participant nr. 25 (2019, Cardiff).

data source underpinning all of that that you could then model the survey on.'[895]

Scottish data sharing participants sketch a substantially similar picture. A policy-maker, working for the Administrative Data Research Centre (intermediary body) in Scotland, explains that introducing the thematic approach to research was one of the key changes from the Administrative Data Research Network (ADRN) to its successor investment, the Administrative Data Research (ADR) UK:

'Under ADRN, you went to the research centre with your project, and if it's accepted it came back to be implemented as a project, but nothing is working like that under ADR UK. Under ADR UK, we need to specify the projects first, and get funding for those, and the projects are run by the researchers who have been brought on board across Scotland.'[896]

As the interviewee explains, this has influenced the overall function of the ADRC in Scotland. To demonstrate consistency with the research themes that are of interest to the Scottish government, the ADRC tries to 'shoehorn things about policy impact' in their recruitment of research fellows, as well as regularly 'run seminars on how to engage with policy makers' for their academics.[897] They express the aspiration that adherence to the thematic approach will enable the ADRC to engage with key individuals in government who will then be integral in shifting the culture around research data sharing through their leadership and their capacity to commit the government to introduce changes:

'We want to get champions within the Scottish government, because we are so far from having a level of awareness there. If we have champions and if we have meetings where many policy-makers attend, then, it's not policy impact, but it's policy interest.'[898]

[895] *Ibid.*

[896] Interview with participant nr. 31 (2020, virtual).

[897] *Ibid.*

[898] *Ibid.*

In the interviewee's view, 'policy interest' is to be defined in a more open-ended and facilitative for researchers manner compared to the more demanding and measurable in a quantitative way notion of 'policy impact'. Another interviewee, who has worked in a consulting role for the ADRC (intermediary body) in Scotland, expresses relief over this change and the introduction of the thematic approach. They welcome the Scottish government's 'greater role in terms of data acquisition and research focus', stressing that it was a 'real struggle' when researchers asked the ADRC to establish on its own initiative a data sharing procedure with the government.[899]

Brigading research projects around specific themes that are of policy interest is, to use the words of a policy-maker working for the Economic and Social Research Council (intermediary body), an effective way to 'steer researchers towards the big questions that the government departments would have'.[900] The perception of policy-relevant research as more appropriate, however, would not only be satisfied with a delineation of the research topic. To make sure that research addresses their worries and interests appropriately, custodians have sought to co-design the research projects that they would provide data to.

iii. Policy-relevance within data sharing frameworks: co-designing research projects

Research 'co-design' or 'co-production' has been a buzzword in the social-science academic discourse for some time.[901] It involves engagement with real-world stakeholders, about whom the research is, as part of the research project activities.[902] Research evaluation frameworks, like the UK's Research Excellence Framework, recognise 'co-production' as a way of demonstrating the 'real-world impact' of research.[903] Data custodians are not blind to the

[899] Interview with participant nr. 34 (2020, virtual).

[900] Interview with participant nr. 19 (2018, Swindon).

[901] T Zamenopoulos and K Alexiou, 'Co-design As Collaborative Research' (2018) (Connected Communities Foundation Series, Bristol) <http://oro.open.ac.uk/58301/1/Co-Design_CCFoundationSeries_PUBLISHED.pdf>.

[902] R White and H van Koten, 'Co-designing for Sustainability: Strategizing Community Carbon Emission Reduction Through Socio-ecological Innovation' (2016) 19(1) The Design Journal 25.

[903] G Chowdhury et al, 'Measuring the Impact of Research: Lessons from the UK's Research Excellence Framework 2014' (2016) 11(6) PLoS ONE.

appeal of impact to researchers and often invoke it to enshrine a right to co-determine the research questions and the methodology used by the researchers within their data sharing frameworks.

The Department for Work and Pensions (DWP) (custodian body) is a prime example of a data custodian who seeks to co-design research projects before allowing access to its data. The department's input may, in some cases, go as far as minimising the researchers' own contribution to the substantive content of the project. An interviewee, who has worked in the past for the Office for National Statistics (ONS) (custodian body) and is currently in academia, elaborates on the process of co-designing a research project with both the ONS and the DWP. Worried that they will struggle to access administrative data, they explain that they first approached the ONS and the DWP before having a concrete idea about their research questions.[904] After getting the buy-in of the departments, the researchers established a 'government stakeholder group', essentially deferring the research questions to the data custodians:

> *'What we're going to do is develop this and have the government departments tell us where they want to go (…) sometimes you're concentrating too much on the academic goals, and the small-level - 'wouldn't this be really interesting as a research project?', whereas we started by saying 'what are you interested in? what do you need to know? what is your constraint?'*[905]

The interviewee describes this experience of getting feedback from the data custodians as 'really useful', even if they concede that they now have to 'put in place a clearer goal':

> *'The thing is we got all these government departments to sign in to this, but we didn't actually think very much about the research goals. We said, basically, help us support this, we are going to put all of this data together and then we can do all these fantastic things. We haven't actually said we are going to do a particular thing! We hadn't thought about that!'*[906]

[904] Interview with participant nr. 42 (2020, virtual).

[905] *Ibid.*

[906] *Ibid.*

An administrative data researcher, who has also negotiated access to DWP data, reports another radical intervention to the substance of the proposed research. More specifically, the DWP convinced the researchers to redesign the substantive focus of the project, making the health impact of benefit sanctions a central part of the research questions instead of the original focus, which was on educational outcomes for children whose parents had received sanctions. The interviewee confirms that this 'was not the original focus we were interested in',[907] yet admits that changing their original focus was required to align their interests with those of the department. They downplay the importance of this change by mentioning that health impact 'was part of the wider research programme, it was ok'.[908] Nonetheless, the promises of smooth data access have not been fulfilled in this particular case, with the researchers still waiting for the necessary permissions by the department for over a year:

> '*It's a set of research questions that DWP analysts themselves have committed to address, but, somehow, I don't know what happened, we are still waiting for their support. It was supposed to be a quick win...*'[909]

Another interviewee, who works as a data analyst for an independent research organisation, concedes that they often have 'no say' about the scope of the research questions.[910] Although they claim that they still pursue their independent research aims, they admit that accessing data for an independent project is often difficult, and using 'whatever is available' as part of an agreement with the DWP is common practice and, often, the only option if they are doing administrative data research.[911] A similar experience is recounted by a data analyst who works for the Administrative Data Research Centre (ADRC) (intermediary body) in Northern Ireland. As they explain, in response to custodians' attempts to shape the research questions, researchers have had to 'come back and completely rework their applications'.[912] This would, in some cases, mean that the projects would change substantively:

[907] Interview with participant nr. 38 (2020, virtual).

[908] *Ibid.*

[909] *Ibid.*

[910] Interview with participant nr. 40 (2020, virtual).

[911] *Ibid.*

[912] Interview with participant nr. 14 (2018, Belfast).

> *'Maybe (they would then) answer research questions that aren't as specific as they would have liked. Some researchers (…) have gone back and reframed their proposal as a Northern Ireland Longitudinal study (…) with a view to branch out to a larger, ADRN study.'*[913]

A policy-maker for the ADRC in Northern Ireland, describes a similar state of affairs:

> '*We have experiences of projects being changed and tweaked, because the data provider is like, no, no, you can ask this, but you can't ask that. You can set this hypothesis, but you can't check this one. An independent research will feel, you've changed my whole research question, this is not what I wanted to ask, how is that independent?'*[914]

This is not to say that data custodians will always be available to accommodate researcher requests. In some cases, such requests may not be feasible from a technical perspective. Still, the interviewee implies that it will often be the custodians' policy priorities, rather than objective technical difficulties, behind the rejection of a particular disclosure. Even when data custodians do not go as far as writing the research questions from scratch or making researchers change their focus, they may still influence the construction of the proposed research projects and seek to 'co-design' them. An administrative data researcher, who has negotiated access with the Scottish government, notes that their colleagues have received nudges to 'target this other population or this subset of population', in spite of their research initially selecting different variables.[915] A similar practice in Northern Ireland is to hold meetings between the researchers and the data custodians, where the latter get to advise the former on 'issues of disclosure and data security (…) as well as on what they would be interested in in terms of policy impact'.[916]

Coming back to Scotland, another administrative data researcher, who has negotiated access with Police Scotland, reflects on their experience. They refer to a project, which originated in

[913] *Ibid.*

[914] Interview with participant nr. 10 (2018, Belfast).

[915] Interview with participant nr. 29 (2020, phone).

[916] Interview with participant nr. 14 (2018, Belfast).

a ‘roundtable meeting’ of academics with data custodians from the Scottish National Crime Agency and Police Scotland (custodian bodies). The interviewee acknowledges that the custodians respected the researchers’ independence, but there were indirect ways of steering the researchers towards their favourable way of designing the project. For example, they would be less enthusiastic and keen on making things move forward unless the researchers would come closer to their view about the way the research would be undertaken. When the researchers remained firm to their views, the project came to a standstill:

> *‘I think this project was not quite what they wanted, and not quite what we wanted, but it would make us both happy enough for people to share data. And, in the end, it all sort of fell apart because we couldn't agree what specifically our research questions were.’*[917]

Speaking about the same data custodian, i.e. Police Scotland, another interviewee, who works as a policy-maker for the Administrative Data Research Centre (ADRC) (intermediary body) in Scotland, reports similar difficulties. They mention that one of the administrative data researchers, who is affiliated with the ADRC, tried to engage with them over a long time to access their data. It took some time to reach this agreement, which was conditioned upon the data custodian’s right to ‘co-produce’ the research:

> *‘There have been 35 meetings over 4 years and the result is that now Police Scotland are going to work with us on the co-production of projects (…) At this moment, as we speak, they are setting up a couple of cautious projects to test the water.’*[918]

The same interviewee mentions that the Ministry of Defence (custodian body) was also keen to co-design with the Scottish ADRC a project using veteran data. In this case, it was not only the data custodians who participated in this process, but also other stakeholders and ‘representatives of the veterans’ community’.[919] The ADRC policy-maker explains that after agreeing on some ‘very simple questions’ with the custodians and the stakeholders, it will now

[917] Interview with participant nr. 35 (2020, virtual).

[918] Interview with participant nr. 31 (2020, virtual).

[919] *Ibid.*

be for the researchers to make the questions 'more academic' so as to satisfy 'fundamental academic rigour' for publication purposes.[920] In other words, instead of starting with an independent research question and then seeking to clarify its policy impact, administrative data researchers often have to start with a policy question and then seek to clarify its interest as an academic research question and the scientific rigour they should apply to its study to meet academic standards.

Like perceptions of administrative data ownership and their related entitlement attitudes, conventional assumptions about the policy relevance of administrative data research effectively filter access to data for independent researchers. Either in an articulated manner, or in a non-articulated manner, the 'rules of the game' often involve the right of data custodians to brigade research requests together or actively intervene in the actual design of the research. Researchers can try to oppose this or to bargain their way out of too many interventions to their research project. Nonetheless, if they want to get data access, researchers may eventually have to respect custodian interventions as a legitimate aspect of the research data sharing procedure.

C. Conclusion: The Salience of a 'Research Data Sharing Culture' for Regulating Decisions

In the UK public sector, access to data for researchers is a privilege, not a right, and as such, it is subject to the unwritten, cultural forces which were presented in this chapter. Although there is a sophisticated legal framework governing access to public-sector data by independent researchers, presented in the previous chapter, the 'rules of the game' that effectively regulate this conduct are non-legal and non-codified. As it has been shown, such rules are not necessarily the outcome of democratic processes, as is normally the case with legislative law-making, but are often produced by the organisations who bargain and clash with one another in the day-to-day workings of the research data sharing space.

Data custodians take advantage of the interpretative latitude afforded to them by a legal framework which does not regulate either administrative data ownership, or administrative data access in a prescriptive manner, filling the gaps in the way they see fit. Not all data custodians subscribe to the same conventions and assumptions, as this chapter has demonstrated, nor do

[920] *Ibid.*

their data sharing practices form a uniform response to the novel challenge of utilising government data to create evidence-based research. Nonetheless, the chapter has presented how their cultural approaches to research data sharing are very effective in shaping the way in which they will create and implement data sharing frameworks. These cultural approaches may be interlinked: e.g. it is reasonable to assume that custodians with strong perceptions of data ownership are also likely to demand clear relevance of research to their policy priorities, resisting accountability and exercising sovereign rights to shape research questions.[921] In any case, the chapter has presented the numerous ways in which the constituents of the 'research data sharing culture' are associated with the formation of specific research data sharing frameworks, enhancing or confining the powers of specific constituents of the presently analysed space.

Having presented empirical findings on the prevalent influences on the formation of the 'rules of the game' in the research data sharing space, this thesis turns to the third and final dimension of the space: the *interactions* between the space's constituents. This and previous chapters have demonstrated that changes in this space are often a product of informal interactions, under the influence of conventional norms and customary assumptions about the activity of research data sharing. Relationships and communications between different actors are seemingly integral to influencing developments, leveraging resources and asserting authority. The following chapter addresses the question of interactions between the actors of this space, enquiring into the ways in which they influence the development of research data sharing cultures, thus shaping the 'rules of the game' that filter access to administrative data.

[921] Further research would certainly shed more light on whether this is indeed the case, see *infra* chapter 8, C.

7. THE CENTRALITY OF MICRO-LEVEL INTERACTIONS IN SHAPING THE REGULATION OF RESEARCH DATA SHARING DECISIONS

> *'This is a definite weakness (…) people who acquire the data are divorced from the uses of the data. Sometimes custodians have research questions and we might have been able to say - if you give us the data, we will use the data for this purpose and you will see these benefits.'* [Interviewee 25]

> *'The political situation at the time was also more sensitive. That naturally makes any minister more cautious. On occasion, getting stuff out was less about 'is this good research?' and more about 'what's the politics going on now?'. Nothing is unaffected by the political side.'* [Interviewee 33]

> *'I mentioned the informal there and I think that actually was crucial in moving forward the work. Lots of things just happened to align (…) Through individuals (…), we were really able to start moving forward, thinking how we would enable the data to be made available for linkage and later for analysis.'* [Interviewee 34]

Previous chapters discussed how organisational actors have been shaping the formation of the 'rules of the game' that filter access to administrative data in the research data sharing space. Particular actors have been triggering the 'opening up' or the 'closing down' of the boundaries of the research data sharing space, either to widen access to administrative data for independent researchers or to restrict the same boundaries and confine data access to certain types of research under certain strict conditions. In this process, it has been shown that a collective 'research data sharing culture' has been the most salient force that shapes the space and, effectively, regulates data access decisions by structuring the terms of reference within which such decisions are taken.[922]

Either with regard to 'opening up' or 'closing down' the boundaries of the space, so-far analysis has suggested that the research data sharing culture has often been reflected in the

[922] See *supra* chapter 2, A, I on the clarification that the interest here is in the process, not the outcomes of decision-making.

interactions between data custodians, intermediary bodies and researchers. In other words, actors have been invoking certain narratives or promoting their assumptions about the appropriateness of research data sharing through their communications with one another. How do these interactions, then, influence the process of shaping the 'rules of the game' in the research data sharing space? This chapter addresses the last, yet crucial analytical dimension of the space:[923] it enquires into the interactions between the different groups operating in the research data sharing space and representing the interests of different organisations. Data from the four UK countries, involving fourty-two (42) data sharing participants working in a variety of roles and for a variety of organisations, shed light on the nuanced ways in which these interactions shape the space.

As it will be argued, while high-level policy makers are, in theory, the ultimate decision makers, commonplace, routine and low-level communications between *data analysts* are more crucial in determining the scope and content of research data sharing frameworks. Disagreements, concerns and crises are negotiated and perceptions about the desirability of administrative data research are shifted through *connections, informal communications and small-scale iterations*, not through high-level, momentous policy decisions. The chapter proceeds from unpacking the role of data analysts as the prominent network behind these interactions, before discussing the nature of the interactions themselves.

A. The Multi-Faceted, Central Role of Data Analysts in the Research Data Sharing Space

In terms of organisational hierarchy within a data custodian body, *analysts* are the lowest-ranked officials in the process of research data sharing. They perform technical tasks and facilitate the actual execution of data disclosure, yet they do not decide about whether such a disclosure shall happen or not. *Policy-makers* are the ones who weigh the risks and the benefits related to a proposed data share and ultimately give the green light, provided that they have received the assurances from *information governance officials* that the disclosure is consistent with legal and other governance (e.g. ethical) requirements. Perhaps understandably, then, data analysts are not celebrated enough in the public-sector data sharing literature. The academic

[923] See *supra* chapter 2, B-D on actors, institutions and interactions as the three dimensions.

literature,[924] as well as reviews of the data sharing practices of public bodies in the UK by the government or by civil society bodies,[925] mainly discuss the role of higher-level officials within data custodians and other data sharing participants. Data analysts are often perceived as merely the technical executors of the policy directions around enabling or constraining research data sharing.

Yet, analysts are the ones who actually manage the datasets, as well as the ones who are well-placed to identify what needs to happen before data access can be granted to external researchers. They are also the ones who normally represent data custodians in their relationships with requesting researchers and intermediary bodies. These relationships often involve the participation of multiple parties and necessitate long-term engagement and constructive collaboration. Analysts, it will be shown, are crucial 'nodes' of the network of interactions that shape the boundaries of the research data sharing space due to their expertise: being administrative data experts, they are uniquely placed to streamline processes by solving problems that are not even visible to other actors. Often coming from a data science or similar research background themselves, they are also keen on enabling more data to be used by researchers. The crucial role of data analysts as mediators within research data sharing interactions is contrasted here with the more high-level, detached role of the two other main groups, i.e. policy-makers and information governance officials.

i. Analysts as experts: knowing the data, shaping the process

Do all analysts perform the same type of duties? An interviewee, who has worked as a data analyst for the Department for Education (custodian body), provides some helpful background on this question. According to them, there are 'three different types of data people':

> *'Those who are responsible for the data collection, those that put out the basic headline statistics for publication and then those who are working with policy specialists to develop how the government operates.'*[926]

[924] E.g. Laurie and Stevens (n 7); Jackson (n 308); Taylor (n 67).

[925] Law Commission (n 2); Thomas and Walport (n 2); REFORM (n 115).

[926] Interview with participant nr. 33 (2020, virtual).

This third type of 'policy-facing analysts' are the ones who are mainly of interest within the present study, since they are able to evaluate whether a dataset can be used for research purposes.[927] To provide some background on the pivotal role of data analysts in the research data sharing space, one needs to associate their authority with their expertise. There are two reasons for which their expertise makes them necessary to making things happen in the context of research data sharing. First, analysts are the only ones who understand the data in respect of the contents of administrative datasets. As the technical experts who have been tasked by public bodies with collecting and curating the datasets that relate to an organisation's functions, they know the substance of the data and, thus, its utility for research. Second, analysts possess the necessary statistical and methodological knowledge to envision solutions to limitations or barriers that arise in the course of a data sharing relationships.[928] Both skills are necessary in creating and implementing research data sharing frameworks.

Although not all data analysts are statisticians,[929] most data analysts share some fundamental knowledge about statistical research and advanced quantitative methods. For example, most analysts are coming from studying degrees with some, often significant, mathematical components, as well as normally have experienced in using data for research or analysis. An information governance official, who works for the Office for National Statistics (ONS) (custodian body), explains that the Office requires a particular level of competence to hire someone as a data analyst: an 'undergraduate degree with a significant component of mathematics, stats, economics, something numeric effectively',[930] or, alternatively three years of research experience using quantitative data.[931] An interviewee, who has worked as a data analyst for the ONS, claims that the UK government is considering establishing a distinct 'track of professional development' for government data analysts, to boost the government's

[927] *Ibid.*

[928] Also see *supra* chapter 4, A, ii-iii on the path-opening role of statisticians and statistical researchers towards formalisation in the research data sharing space.

[929] E.g. at the electronic Data Research and Information Service Scotland, interview with participant nr. 36 (2020, virtual): 'We've got a bit of a mix, we've got people on the team who have come through analytical roles in the NHS (…) We've got a couple of researchers (…) We've got people who were managing data rather than IT within the NHS'.

[930] Interview with participant nr. 1 (2017, Titchfield).

[931] Office for National Statistics (n 148).

expertise in that field.[932] A policy-maker, working for the ADRC in Scotland, welcomes this idea, observing that the lack of statistical-methodological expertise is one of the main reasons behind the inability of Scottish local authorities 'to benefit from the data they collect'.[933] Considering the technical complexity of administrative data research,[934] such expertise is integral to envisage solutions to data sharing challenges, regardless of the subject matter that the data's contents relate to.

Because of their expertise, data analysts are playing a central role in four (4) key moments of the research data sharing lifecycle:[935] *identifying* the types of data that are of interest for research purposes; *establishing consistent standards* in terms of organising and classifying datasets and variables within datasets to enable data linkages between different organisations; *cleaning and curating* the datasets so that they can be used by researchers and *establishing safeguards* to reconcile the need for more data that will allow engagement with more or broader research questions with data protection and minimisation requirements. The functions that the data custodian bodies exercise within these four key moments are, to a significant extent, carried out by data analysts, even if the latter are not formally seen as making the final decisions.

To start with the *identification* of the appropriate administrative dataset for research purposes, this might in some cases necessitate the creation of a dataset anew and in other cases the identification of a dataset among existing datasets. In the first case, creating datasets for the first time might necessitate the heavy task, from an administrative perspective, of having to register the content of the dataset for the first time. Even if this process is to an important extent automated,[936] it is still a considerable challenge in terms of understanding the key parameters that need to be satisfied for a high-quality dataset to be built. As an administrative data

[932] Interview with participant nr. 28 (2020, London). This track of professional development would essentially encourage government officials to invest into gaining statistical skills by establishing relevant incentive structures in the public sector.

[933] Interview with participant nr. 31 (2020, virtual).

[934] Administrative Data Taskforce (n 23).

[935] S Aidinlis et al, 'Building a Justice Data Infrastructure: Opportunities and Constraints' (2020) (AI FOR ENGLISH LAW UKRI GRANT) <https://www.law.ox.ac.uk/sites/files/oxlaw/ukri_justice_data_report_fv_0.pdf>.

[936] SAIL Databank (n 605).

researcher with experience in negotiating access to health data in Scotland explains, the role of data analysts is critical because 'everyone wants to harvest databases, but no one wants to input the data'.[937] Data analysts are the ones with the crucial task to ensure good quality of input data, or to 'clean the data up a bit' when the input data are not of ideal quality.[938] Another policy-maker, who has worked for the HM Revenue and Customs (HMRC) (custodian body), identifies internal disparities in ways to register data as a real obstacle in creating a uniform dataset:

> *'There are 30 or 40 different lines of duty in HMRC - stamp duties, tax credits, different things - they all have different data systems and within them there are different strands of recording data developed by different IT contractors.'*[939]

Even if the dataset exists, however, it might still be a challenging task to identify which one of the many datasets that data custodians hold is appropriate for research use. In respect of achieving this, sound knowledge of the strengths and the limitations of the datasets is necessary, particularly in the case of large ministerial departments holding a massive wealth of information.[940] Many interviewees acknowledge the significance of this skill. An interviewee who works for the Office for National Statistics (custodian body) as a policy-maker explains that data analysts are their first port of call when they receive a data sharing request:

> *'They are the experts in the data, they are very much giving advice, guidance to us on the feasibility of a research proposal (…) they would advise for instance that oh, that is the annual population survey that is your most useful dataset for that.'*[941]

Including data analysts in all stages of the decision-making process, as the interviewee explains, is necessary to ensure that any questions that policy-makers might have about a request can be addressed.

---

[937] Interview with participant nr. 39 (2020, virtual).

[938] *Ibid.*

[939] Interview with participant nr. 32 (2020, Oxford).

[940] Law Commission (n 2) 10.8.

[941] Interview with participant nr. 4 (2017, Titchfield).

This exercise of identification is often hampered by the lack of *consistent standards* among datasets and variables, so that data custodians, either within their own organisation or in collaboration with another data custodians, have compatible data resources that can be linked together to produce a valuable analytical resource. In many cases, the lack of such standards has to do with how vast, massive and fragmented in terms of standardisation administrative datasets are. Since such datasets are primarily collected to serve the primary functions of public bodies, they are far from consistently organised to be used by researchers. Data analysts are integral to overcoming this barrier and turning the datasets into consistent with one another resources. A policy-maker, who works for the Department for Work and Pensions (DWP) (custodian body), explains their gratitude for the analysts of their department helping them with making sense of DWP datasets. They highlight that this is an incredibly complex process, as the datasets are 'incredibly messy' and no external user could understand how to use them if they are not first organised in a more standardised way:

> *'There are some datasets where, if we release it tomorrow (…) you just know that anybody would look at it and say 'what on earth is this?' (…) there is that task which should not be underestimated of turning the data which we use every day, which can be quite messy into something which is more readily understood by an external user.'*[942]

Data analysts, possessing sound knowledge about the datasets, are able to 'tie the data together' and provide such consistency.[943] This is corroborated by two interviewees working for the DWP. A policy-maker working for the department explains that they find it impossible to understand what exactly can be done with the data, particularly when there is 'such a wealth of data covering such a range of topics'.[944] According to them, even analysts themselves need to collaborate closely, since 'there isn't anybody who actually understands it all'.[945] This is why the department has established a sophisticated network of 'Lead Analysts', responsible for

[942] Interview with participant nr. 7 (2017, Sheffield).

[943] *Ibid.*

[944] *Ibid.*

[945] *Ibid.*

distinct areas of DWP data, employing 'around 20 of those' at the time of writing.[946] An information governance official working for the department concurs that these expert analysts 'really do know these bits better' and have a 'very deep knowledge of the research that's happened in their area and knowledge gaps in their area'.[947]

The same rationale of relying on data experts to achieve consistency between datasets applies all the more when it comes to linking datasets with the datasets belonging to another custodian. For example, a researcher may be interested in linking health data belonging to the National Health Service (custodian body) with DWP benefit sanctions data to assess the health impact of DWP's enforcement policies.[948] An information governance official, who works for the Honest Broker Service (custodian body) in Northern Ireland, explains that, in such cases, the linkage is facilitated by another organisation (an intermediary body), which provides the secure environment to ensure that identification risks created by the aggregation of data will be minimised.[949] As the interviewee explains, the presence of the data analysts working for this 'trusted' intermediary is necessary in all meetings concerning the provision of access to linked data since they have the 'expertise over the secondary care data'.[950] In a similar vein, an interviewee, who has worked for the HM Revenue and Customs (HMRC) (custodian body) as a policy-maker, recalls how important the contribution of the department's data analysts was in developing consistent data quality and interoperability standards.[951] Had these standards not been established, the whole aspiration of the HMRC to create a datalab that allows access for external researchers would have been put to rest due to financial concerns:

> *'A major obstacle arises around simply tracing and matching between data sets which can dramatically increase cost and expense for data sharing – sometime*

[946] *Ibid.*

[947] Interview with participant nr. 8 (2017, Sheffield).

[948] S Pattaro et al, 'Labour Market and Wider Impacts of Benefit Sanctions: A Scoping Review' (2019) 17th Annual ESPAnet Conference, Stockholm, Sweden, 5-7 Sept 2019.

[949] Harron et al (n 140).

[950] Interview with participant nr. 11 (2018, Belfast).

[951] Aidinlis et al (n 928) 24.

> *terminally. Having common data standards and better data quality addressed this.'*[952]

Having identified and consistently organised the appropriate datasets for research uses, data custodians also need to ensure that the data are *curated* and *cleaned* for use by researchers.[953] While this sounds an extremely technical and, arguably, mundane task, the truth is that it is as necessary a step for a research data sharing initiative as any other. Sharing datasets with irrelevant variables or not updated variables may yield unsatisfactory results, both from the perspective of heightened data protection risks (i.e. lack of conformity with the requirement to process as little data as necessary) and reduced research utility (i.e. using data points that are not updated and, thus, less useful). An administrative data researcher, with experience in the Scottish data sharing landscape, describes data analysts as the ones who 'curate the dataset' and clarifies that these are normally the first point of contact if a researcher is interested in gaining access.[954] As an ex-Department for Education (custodian body) data analyst explains, analysts would be sitting at an internal Approvals Panel within the department to 'ensure that the data have been prepared appropriately' and that the requesting party will be able to make sense of them.[955]

Beyond cleaning and curating datasets, however, the role of data analysts is generally critical in *establishing safeguards* to facilitate researcher access to the data. After the datasets have been prepared for research use, there are still many challenges around ensuring that access strikes a balance between maximising research utility and complying with legal and ethical safeguards.[956] Analysts are well-placed to help custodian bodies to strike this balance by advising on the 'kind of methods' that can be applied to particular datasets in the interest of addressing particular research questions.[957] Thanks to their expertise, they understand a lot about 'statistical disclosure and security', while at the same time ensuring that 'the research

[952] Interview with participant nr. 32 (2020, Oxford); also see *supra* chapter 4 p 102 for the use of the HMRC datalab by external researchers.

[953] SAIL Databank (n 605).

[954] Interview with participant nr. 35 (2020, virtual).

[955] *Ibid.*

[956] Aidinlis et al (n 928) 31.

[957] Interview with participant nr. 14 (2018, Belfast).

questions are still being answered and are not being compromised'.[958] In some cases, analysts may help by performing a particularly challenging statistical technique to enable meaningful data sharing that can allow researchers tackle their research questions without compromising the protection of sensitive data. A policy-maker, working for the Department for Work and Pensions (DWP) (custodian body), recalls such an example:

> *'Potentially, depending on the nature of the work, there may also be a need to do some fuzzy matching of an external dataset against a DWP data. And there is a separate team that does that, because there are very few analysts in DWP who have access to data with names, addresses and all that, so you can do this kind of matching.'*[959]

A data analyst, working for the Knowledge and Analytical Services (custodian body) of the Welsh government, goes as far as implying that the data analysts' recommendations with regard to methods and project particulars shall receive deference from the policy-makers:

> *'You know, a policy maker actually shouldn't know the methods whereby we deliver that evidence. They should tell us what their evidence need is. It's our job to tell them the best way to deliver that.'*[960]

Some interviewees reflect on scenarios where data analysts used their statistical and methodological expertise to break an impasse in research data sharing negotiations by providing assurances that sufficient safeguards have been applied to the procedure. For example, the Health Trusts (custodian bodies) in Wales were concerned with the potential risks of re-identification of individual patients if the SAIL Databank collected huge amounts of health data and linked it with administrative datasets belonging to the Welsh government.[961] It was not a political assurance or a high-level declaration that allowed the SAIL Databank to allay these worries and proceed with establishing their data sharing framework, now considered an example of international best practice. It was the ingenuity of a team of data analysts,

[958] *Ibid.*

[959] Interview with participant nr. 7 (2017, Sheffield).

[960] Interview with participant nr. 26 (2019, Cardiff).

[961] SAIL Databank (n 605).

working for the NHS Wales Informatics Service (intermediary body), as an interviewee who works for SAIL in a policy-making capacity explains:

> *'There were people there who were in a position who could create this algorithm that allowed the de-identification of the data, so that the data, when they were analysed, were only used in de-identified form. (…) Them developing the algorithm was really crucial, actually, for good data quality.'*[962]

A similar function is operated by the data analysts of the National Records of Scotland (NRS) (data custodian) when it comes to linking their data with National Health Service data: the NRS team apply a technique called 'probability matching' to produce unique identifiers as a way of encrypting the datasets before providing access to researchers.[963] Through such techniques, analysts effectively facilitate the compliance of research data sharing with the principles of data minimisation and security in data protection law,[964] ensuring that researchers are only accessing data they strictly need for their projects. It will often be the case that advanced statistical expertise is needed to achieve this:

> *'We would discuss issues like do they want what we call a long or a wide dataset, do they want one row per item (…) and we need to be re-assured that the researcher is comfortable with linking a long dataset to a wide dataset because it is quite a challenging process. You need some good STATA skills.'*[965]

The analysts' central role in establishing data sharing safeguards is acknowledged, in practice, by intermediary bodies or researcher groups which employ themselves data analysts with a view to making communication more efficient with data custodians. As a data analyst who works for the Knowledge and Analytical Services (custodian body) of the Welsh government explains, they do not feel that they can negotiate data access efficiently with people who do not have the required skills to understand the data:

---

[962] Interview with participant nr. 21 (2019, Swansea).

[963] Interview with participant nr. 36 (2020, virtual).

[964] Article 5 of the EU GDPR.

[965] Interview with participant nr. 12 (2018, Belfast). Note that STATA is a statistical software package used by social scientists.

> *'This is a definite weakness (…) people who negotiate are often divorced from the uses of the data. Sometimes we have research questions and requesting parties might have been able to say - if you give us the data, we will use the data for this purpose and you will see these benefits.'*[966]

Some intermediary bodies have, indeed, heeded this call and have hired data analysts to ensure that their requests are communicated to the data custodians in the right manner. The Administrative Data Research Centre (ADRC) in Northern Ireland is a particularly interesting case study in that regard. As an interviewee, who has worked as a policy-maker for the ADRC-Northern Ireland, explains, the Centre has appointed two 'expert administrative data researchers' with data analyst skills as 'Statistical Methodological Officers (SMOs)'.[967] The role of the SMOs is to scope the feasibility of requests, based on their knowledge of the data that are available for research use in the Northern Irish public sector. Another interviewee working for the same body as an information governance official concurs that it is important to use people with statistical expertise when trying to 'establish a good working relationship' with data analysts working for data custodians.[968] As another policy-maker, who works for the Economic and Social Research Council (intermediary body), puts it: 'if you do not know the dataset, how are you supposed to know what to ask for?'.[969]

Using their expertise in managing data, as well as their subject-matter knowledge of particular datasets, analysts have been instrumental in the development of research data sharing processes. Through this role, they have been the ones who have often authoritatively interpreted data sharing criteria and requirements, shaping the formation of the 'rules of the game' in this space. Crucially, even if this is not formally required from them, they have seen themselves as mediators between data custodians and other data sharing participants, seeking to facilitate research data sharing processes in a number of ways.

[966] Interview with participant nr. 25 (2019, Cardiff).

[967] Interview with participant nr. 10 (2018, Belfast).

[968] Interview with participant nr. 9 (2018, Belfast).

[969] Interview with participant nr. 19 (2018, Swindon).

ii. Analysts as mediators: helping researchers, streamlining data sharing processes

Data analysts, while being often under the supervision of higher-level colleagues like information governance officials, are central in the research data sharing space not only because of their expertise, but also because of their tendency to bridge the distance between custodians and other actors.

As mentioned, analysts are the ones who first get to assess a research data access request received by a data custodian. This often means that they are the ones who get to set the terms of the discussion for the remainder of the data access negotiation. An information governance official, working for the Office for National Statistics (custodian body), explains that they supervise a team, consisting of a number of data analysts. These analysts take it upon themselves to clarify the scope and the requirements that the researchers will need to meet for the access to be granted:

> *'The researcher's application comes to my team. We then reflect within the team (…) do you agree that this data is appropriate for this research? Can their research goals be met by using this data?'*[970]

There are various ways in which analysts, both ones working for data custodians and ones working for intermediary bodies seeking to help researchers negotiate data access with custodians, assist researchers in understanding data access requirements and, thus, streamline data sharing processes. A common requirement is that data sharing requests are precise enough to allow custodians to assess whether the requested disclosure conforms with the principles of proportionality and data minimisation.[971] Analysts are the officials who assist researchers in turning a broad interest in some administrative datasets into a concrete request that will satisfy custodian requirements. A data analyst, who works for the Administrative Data Research Centre (intermediary body) in Northern Ireland, describes the main steps of this process. First,

[970] Interview with participant nr. 1 (2017, Titchfield). In that sense, analysts operate as 'regulatory stewards' who can help researchers 'navigate difficult regulatory spaces and improve the overall quality of research', see Dove (n 11) 183.

[971] UK Statistics Authority (n 70).

they receive researcher queries via email and seek to clarify what exactly is the researchers' analytical interest, seeking to determine whether this is 'the right thing for them'.[972] After this is established, they would meet with the researchers and 'talk through all the different processes to access the data',[973] helping them understand how they will satisfy the various governance requirements, either examined directly by the custodians or by other bodies like Research Ethics Committees. When the researchers have prepared the documentation for submission, the analysts would again help refine the application before it is formally considered by the custodian, 'reading through their drafts and pointing out areas where they might want to add a bit more, or take a bit away'.[974]

A colleague of theirs, who has worked as a data analyst for the Administrative Data Research Network (intermediary body), breaks down the ways in which a request could be 'narrowed down' to satisfy custodians:

> *'What we would do, we would have more expertise from the experience of helping them narrow it down to something that didn't involve the whole of the world. Temporal scoping, which years - geographical scoping, e.g. England & Wales, or England & Scotland.'*[975]

When researchers had a 'very vague idea' about the project,[976] the interviewee explains, such scoping would take longer and it would involve long e-mail exchanges and phone conversations. Another data analyst, who has worked for the same intermediary body, admits that they would receive some open-ended questions e.g. 'I would like to do some research with some health data, can you do that for me?'.[977] In these cases, the analysts would probe the researchers to clarify their interest in the data:

[972] Interview with participant nr. 15 (2018, virtual).

[973] *Ibid.*

[974] *Ibid.*

[975] Interview with participant nr. 5 (2017, Essex).

[976] *Ibid.*

[977] Interview with participant nr. 6 (2018, phone).

> *'With those people who have a very vague enquiry, is basically we send them back an awful lot of questions which force them to articulate more clearly what is it they want.'*[978]

In the interviewee's mind, the best mechanism to get the researchers down to concrete questions is the requirement of submitting a comprehensive written request:

> *'It was an 11-page word document, because there is quite a lot of information to catch on that. And basically we would use that as the tool to try and make people clear about what they want.'*[979]

This does not mean that analysts will be strictly requiring this information from the researchers without assisting them in the process of thinking about the difficult questions around data access. It will often be the case that researchers will be using the feedback of data analysts while preparing their application forms. Another interviewee, who works in an information governance capacity for the Electronic Data Research and Innovation Service (intermediary body) in Scotland, notes that they see their role as two-fold: help researchers make their requests more concrete, but also 'educate the decision-makers about the fact that sometimes you don't know what you need until you actually get into the analysis element of it'.[980] An information governance official, working for the Honest Broker Service (HBS) (custodian body) in Northern Ireland, describes their own process of 'triaging' requests that HBS receives by external researchers. They explain that they oversee a team of analysts who first clarify with the researchers crucial parameters like 'variables' and the 'sample sizes', enabling them to engage in a 'preliminary exploration on the datasets, in terms of quality, quantity of the data and the feasibility'.[981] They stress the significance of this initial scoping exercise for the data sharing process to proceed:

> *'So, we (in the team with the analysts) have our discussion, an in-depth discussion, about what their intentions are, what the aims of the project are,*

[978] *Ibid.*

[979] *Ibid.*

[980] Interview with participant nr. 36 (2020, virtual).

[981] Interview with participant nr. 11 (2018, Belfast).

> *perhaps tease out a bit more the variables they are interested in, sometimes they have only given a broad name and indication of what it is they are looking at. We need to tease out the actual variable name as it is coded and held on the database.'*[982]

After this initial stage, the HBS analysts, like their counterparts in other bodies, assist the researcher in 'drafting the application',[983] being themselves familiar with the way in which the HBS governance board interpret the custodian requirements. The mediating role of data analysts is particularly highlighted in this case, since data analysts are also part of the data custodian organisation and provide a helpful understanding of the organisational requirements to the researchers. A policy-maker, working for the Department for Work and Pensions (DWP) (custodian body), further explains the importance of this mediating role played by the analysts not only in respect of helping researchers, but also in terms of helping their own colleagues within a custodian organisation. This actually enhances the mediating influence of the analysts, since both sides of the negotiation rely on their insights to make sense of the process. As the policy-making interviewee from DWP claims, they ask data analysts to refine the requests 'into reasonable shape' before passing them on to their policy colleagues, expressing dissatisfaction at their previous lack of refinement:

> *'Some of the requests which come through are sometimes very vague, sometimes they ask for a huge amount of detail, it's not obvious why they need that amount of detail for all the research which they propose.'*[984]

Analysts do not exhaust their mediating role in helping researchers with properly articulating their requests. Crucially, analysts tend to see themselves as close to researchers and often speak on behalf of researchers to their colleagues within custodian bodies. A data analyst, who has worked for the Administrative Data Research Network (intermediary body), attributes this to their background as data managers and users:

---

[982] *Ibid.*

[983] *Ibid.*

[984] Interview with participant nr. 7 (2017, Sheffield).

> *'You get more of the appreciation of the ability to get things done from the analysts, the data using people, (whereas) the people dealing with information governance (…) are more in the kind of defensive, cautious way (…) The analysts might give you interesting insight that I'd like to see, how to we go and make it work the right way.'*[985]

The background of analysts is contrasted here with the background of information governance officials who are more interested in limiting data uses for legal compliance and risk management reasons. An interviewee, who works in an information governance capacity for the Northern Irish Statistics and Research Agency (custodian body) explains that data analysts form an informal network across the Northern Irish public sector. Even if they are not the ultimate decision-makers, they often provide crucial 'buy-in' for a research data sharing process to be established,[986] by convincing their policy-making colleagues about the merits of investing resources into that. A data analyst, working for the Department for the Economy in Northern Ireland (custodian body), provides some background on this aspect of their role:

> *'A request would come to the analytical services, to ourselves, and (…) I myself am very open to doing an ADRC project. So, I think scepticism may be overcome (…) I would recommend it to policy folks if we felt there is sufficient benefit to it, sufficient merit (…) The Department would heavily rely on advice from us in terms of whether to share data.'*[987]

An information governance official, who works for the the Electronic Data Research and Innovation Service (eDRIS) (intermediary body) in Scotland describes how their team of data analysts represent the interests of researchers in negotiating access to Scottish government data. eDRIS analysts mediate between researchers and analysts working for the custodians, ensuring that the latter have sufficient explanations about the requested data and the information governance concerns that they may have, 'avoiding questions going back to the applicants'.[988] Many researchers who have negotiated access to data in Scotland highlight the critical

[985] Interview with participant nr. 6 (2018, phone).

[986] Interview with participant nr. 9 (2018, Belfast).

[987] Interview with participant nr. 16 (2018, phone).

[988] Interview with participant nr. 36 (2020, virtual).

mediating role of eDRIS analysts. An administrative data researcher, with experience in negotiating access in the Scottish public sector, hails eDRIS analysts for their sound understanding of governance issues like statistical disclosure control, expressing relief for being represented by them in access negotiations.[989] Another researcher, with experience in linking health with non-health data in Scotland, concurs that eDRIS analysts 'sort things out for you' and clarify things with the applicants in case they see problems in the application.[990] In a similar vein, a researcher, with experience in linking census data with health data in Scotland, expresses their admiration for eDRIS analysts:

> *'I'm absolutely beholden to my eDRIS co-ordinator who is incredible, I don't know how he does it. It's ridiculously messy.'*[991]

What makes analysts so motivated to act as liaisons between custodians and researchers to help things move forward? While they are not necessarily researchers themselves, their knowledge about the datasets and their statistical expertise makes them more sympathetic to research uses of administrative data. Due to this sympathy, in essence, analysts often aspire to broaden research data sharing frameworks, convincing custodians to make more of their data available for research use. An interviewee, who has worked for the Administrative Data Research Centre (intermediary body) in Northern Ireland in a policy-making capacity, describes that their analysts perceive 'challenging the data providers along the lines of, well, why can we not have that?' as part of their remit.[992] An analyst colleague of theirs, indeed, takes great pride in pressuring the data custodians of the Northern Irish Statistics and Research Agency (NISRA) (custodian body) to make more data available for research:

> *'I have spent a long time giving the NISRA guys hard time for why anything hadn't happened. (…) I would sort of argue back and say ''look, you provided it here, you provided it there'' and we would have an argument. Eventually, we would have a deal, somewhere in the middle usually.'*[993]

[989] Interview with participant nr. 37 (2020, virtual).

[990] Interview with participant nr. 39 (2020, virtual).

[991] Interview with participant nr. 34 (2020, virtual).

[992] Interview with participant nr. 10 (2018, Belfast).

[993] Interview with participant nr. 15 (2018, virtual).

Another way in which analysts contribute to building bridges between researchers and custodians is by training researchers to use data safely and appropriately, helping them understand the existing best practices in data management. A data analyst, who has worked for the Administrative Data Research Network (intermediary body), reflects on their contributions to developing a bespoke 'safe researcher' training for applicants who were interested in pursuing research through the ADRN:

> *'And we developed the training, the bit where the researchers come and have a day or half day session, teaching - learning sessions, when they are taught on ways to handle data, ways that they can make sure the data they give to the staff at the secure environment does not contain personal information.'*[994]

As the interviewee notes, this training was inspired by the Office for National Statistics (ONS) (custodian body) 'approved researcher' training. A data analyst, who has worked for the ONS in the past, clarifies that this training framework, created by analysts, combined methodological knowledge around statistical analysis of administrative data and guidelines on how to conform with legal requirements and information governance standards.[995] Another analyst, working for the Administrative Data Research Centre (intermediary body) in Northern Ireland, adds that they provide training sessions for researchers interested in improving their skills in software packages used for administrative data research. They also offered 'a number of accreditation courses on data security and statistical disclosure control'.[996]

Through their expertise in the data and in statistical analysis, analysts perform crucial mediating roles in the research data sharing space. Nonetheless, one might think that their crucial role is still limited by the fact that higher-ranked officials, such as information governance managers or policy-makers, hold the ultimate decision-making authority. Analysts do not seem to hold any decision-making competences. Is this, indeed, the case?

[994] Interview with participant nr. 5 (2017, Essex).

[995] Interview with participant nr. 28 (2020, London).

[996] Interview with participant nr. 14 (2018, Belfast).

iii. Analysts as decision-makers: blurring the lines between 'recommending' and 'deciding'

As it was previously mentioned,[997] analysts are the lowest-ranked officials in the organisational hierarchy of the decision-making space. Yet, their expertise and knowledge about the data, as well as their mediating role in building bridges between data custodians and requesting parties, effectively blurs the boundaries between the notions of 'recommending' and 'deciding' when it comes to analysts' assessments.

For policy-makers in data custodian bodies, it is a matter of authority to reserve for themselves the last word on whether a research data access request will be granted. It is a common perception among participants in the research data sharing space that 'you need to have someone at the top saying ''make this happen''.[998] A data analyst, who has worked for the Department for Education (DfE) (custodian body), explains that the power to establish and authorise a data sharing procedure belongs to the policy-makers, ultimately leading up to the highest level of authority within the civil service, i.e. ministers:

> *'The policy team specify the policies that they are developing and how this will work - they would then come to my team and say we want to collect this data (…) In terms of making the data available, the ministers would have their say in that (…) They are the ones who have the power to make decisions*.'[999]

A policy-maker, who has worked for the HM Revenue and Customs (custodian body), concedes that ministerial decisions can be 'a bit more complicated, a bit more whimsical perhaps'.[1000] An information governance official, working for the Department for Work and Pensions (DWP) (custodian body), caveats this by pointing out that there are 'certain protocols referring to decisions made by ministers'.[1001] In practice, as an interviewee who sits at the Confidentiality Advisory Group (intermediary body) explains, many of these decisions will be

[997] *Supra* chapter 3, A.

[998] Interview with participant nr. 37 (2020, virtual).

[999] Interview with participant nr. 30 (2020, virtual); it is important to clarify that this is only applicable to ministerial departments, since other public bodies do not have ministers.

[1000] Interview with participant nr. 32 (2020, Oxford).

[1001] Interview with participant nr. 8 (2017, Sheffield).

'delegated to a senior civil servant' by the ministers, unless there is some 'political sensitivity to the decision'.[1002] The DWP information governance official confirms that it is their intention too to have as much data sharing decisions as possible delegated from the ministers to other officials:

> *'Ministers are still getting a feel for the benefits and the costs and risks. As this develops, what we would hope/expect to see is that Ministers go: ''I am fine with all the issues, you guys know what you are doing''. Potentially at some point they might go: ''don't bother me with these anymore''.'*[1003]

In any case, as a data analyst working for the Department for the Economy in Northern Ireland (custodian body) stresses, there is a clear distinction between analysts and the 'policy side of things' who are 'independent' from the analysts.[1004] In other words, there is a clear division of roles between analysts and policy-makers, with the latter having the final word on data access decisions. Another data analyst, who has worked for the DfE, confirms that policy-makers had the ultimate authority with regard to research data sharing. They confess that the policy people even had the authority to block an internal research project if it was 'politically difficult' for the government of the day:

> *'If you did stuff that was difficult for the government, it was unlikely to get published (…) the political situation at the time was also more sensitive. That naturally makes any minister more cautious. On occasion, getting stuff out was less about ''is this good research?'' and more about ''what's the politics going on now?''. Nothing is unaffected by the political side.'*[1005]

A policy-maker, who works for the Administrative Data Research Centre (intermediary body) in Scotland, similarly reflects on the difficulties created by the fact that policy-makers are the ultimate decision-makers in research data sharing. They lament the lack of patience and more strategic vision on behalf of policy makers:

[1002] Interview with participant nr. 27 (2019, phone).

[1003] Interview with participant nr. 8 (2017, Sheffield).

[1004] Interview with participant nr. 16 (2018, phone).

[1005] Interview with participant nr. 33 (2020, virtual).

> *'They usually want an answer to something that week, or that month. My feeling is they are interested in quick answers to solve their particular issue that week or month, than actually engaging with a body of evidence that will help in the medium to long term evidence-based policy making across the board.'*[1006]

A closer look at the micro-level dynamics of the space, however, somewhat nuances the picture of the almighty policy-maker. While formally not holding decision-making powers, analysts do a lot of the work in refining and organising data sharing processes, effectively framing the way in which higher-ranked officials exercise their competences. The understanding of such officials is dependent, to a large extent, on the groundwork laid out by the analysts. This results in analysts becoming much more than simple advisers. To a data analyst, who has worked for the Office for National Statistics (ONS) (custodian body) in the past, this is due to the fact that policy-makers are detached from the data sharing process, whereas data analysts are on top of it:

> *'Data access forms are usually designed by policy people who are a long way removed from the process (…) what you want is something that people can read and it makes sense to them and they can agree to.'*[1007]

An administrative data researcher, who has negotiated access with Scottish data custodians, corroborates that policy-makers are often unhelpful due to them missing the detail of research data sharing like the 'specifics of the dataset'.[1008] They recall a paradigmatic incident where they had to negotiate with policy-makers who were unclear about the specifics and actually misled them about their possibilities to proceed with data sharing:

> *'We were told that you have to start at the (redacted) point of the system and you can't have before that. And then we start talking to the guy who would look after the big database and he said well that's not true, you can actually do that*

[1006] Interview with participant nr. 31 (2020, virtual).

[1007] Interview with participant nr. 42 (2020, virtual).

[1008] Interview with participant nr. 35 (2020, virtual).

> *if you manually do this and give me some set parameters, I'll be able to stitch that dataset for you. That kind of gap between the different people.'*[1009]

The same interviewee adds that senior, policy-making officials are also struggling to stay on top of data sharing specifics because they simply have 'other kind of priorities' and 'more things to do than facilitate an academic project'.[1010] This is shared by a data analyst who has worked for the DfE in the past. Policy makers are the ones with the competence to decide, yet they often 'don't have the time to invest in the level of detail'.[1011] As another analyst, who works for the Department for the Economy (custodian body) in Northern Ireland, explains, this often results in policy-makers 'heavily relying' on data analysts when making up their mind on whether to share data.[1012] Another interviewee, who has worked for the ONS in an analyst capacity in the past, explains that they were once able to convince some of their policy-making colleagues to delegate significant data sharing competences, like the statistical disclosure control,[1013] invoking their expertise in the data:

> *'We said to the (redacted) team, why don't we do this disclosure control? We could make these decisions quickly - we'll send you what we've done, our decisions, and you can check it. You can keep an eye on this, but, basically, we will do the work (…) no one objected to that and no one thought that was a bad plan.'*[1014]

Interestingly enough, another reason for delegating data sharing competences to analysts relates to political sensitivity, especially in the case of policy-making departments. When a data access decision may become politicised, e.g. rejecting access to a project whose research questions can be interpreted as critical to the government of the day, it is more palatable for

---

[1009] *Ibid*.

[1010] *Ibid*.

[1011] Interview with participant nr. 30 (2020, virtual).

[1012] Interview with participant nr. 16 (2018, phone).

[1013] Statistical disclosure control is a 'a technique used in data-driven research to ensure no person or organization is identifiable from the results of an analysis of survey or administrative data, or in the release of microdata', see C Skinner, 'Statistical Disclosure Control for Survey Data' (2009) <http://personal.lse.ac.uk/skinnecj/SDC%20for%20survey%20data%20S3RI.pdf>.

[1014] Interview with participant nr. 42 (2020, virtual).

policy-makers to keep themselves out of the decision. The rejecting decision can, then, be portrayed as a technical decisions by technical stuff, i.e. the analysts. As an information governance official who works for the ONS explains, policy makers 'don't like having to be the person who gets to say no publicly'.[1015] Essentially, a policy refusal of a request turns a supposedly technical procedure of data sharing into a political matter of endorsing or rejecting particular types of research agendas. A data analyst, who has worked for the Administrative Data Research Network (intermediary body), confirms that, reflecting on some feedback they were given by their data analyst counterparts who work for the DWP. In their words, policy people 'never really want to find themselves in the position where they were asking their higher up, i.e. the minister, to refuse to do something'.[1016] It would be preferable for them to have analysts responsible, or at least nominally responsible, for some aspects of the data sharing process.

Furthermore, analysts are much more centrally involved in the practicalities of data sharing compared to information governance officials who specialise in legal compliance and risk management of a data access request. Like policy-makers, information governance officials hold significant decision-making competences, but their role is practically translated into a compliance check that takes place *after* most of the access negotiation has taken place, unless a request is 'clearly illegal'.[1017] An administrative data researcher, with experience in negotiating access with Scottish data custodians, claim that they have hardly reached the stage of liaising with information governance officials:

> *'Yeah, I guess, I was thinking about this question and I think it's never really come up, if you see what I mean. We never really got to the legal point in that process, it's always been trying to get people to agree - is this feasible and do you want to do this? Rather than, is this legal?'*[1018]

A data analyst, who has worked for the Administrative Data Research Network (intermediary body), corroborates this picture of a separate and subsequent information governance

[1015] Interview with participant nr. 1 (2017, Titchfield).

[1016] Interview with participant nr. 6 (2018, phone).

[1017] Interview with participant nr. 5 (2017, Essex).

[1018] Interview with participant nr. 35 (2020, virtual).

participation in the data access negotiations. They recall that information governance officials of the custodians were the only ones responsible for determining the lawfulness of the disclosure, with the requesting and intermediary parties often accepting the custodian determination.[1019] A policy-maker, working for the DWP, adds that the information governance side is entirely distinct from the policy weighing of the merits and the risks related to an access request. Once policy makers have 'considered all of these requests and decide which ones have merit to pass on',[1020] the request will be considered by the information governance officials who will opine on 'whether it is legal to do or not'.[1021] A data analyst, who has worked for the DfE, explains that information governance officials do not have a more central role in designing a data sharing framework because their job is to 'reduce risk about making data available', not maximising the 'value of granting access to data'.[1022]

As it has been shown, the humble group of data analysts plays a multi-faceted and central role in the research data sharing space. Interactions between analysts representing different actors, i.e. data custodians and intermediary bodies or researchers, as well as interactions between analysts and policy-makers or information governance officials within data custodians, matter. Through these interactions, analysts are able to leverage their expertise in the data and in statistical analysis to build bridges between different parties and offer authoritative recommendations. In doing so, they shape research data sharing processes and, through that, the content of data sharing frameworks, often with a view to facilitating research data sharing. Higher-ranked officials, like policy-makers and information governance officials, often find themselves implicitly or explicitly reliant on the understanding of research data sharing processes that is offered to them by analysts. To complement this finding about low- rather than high-level officials setting the boundaries of the research data sharing space, I now turn to the nature of the interactions between the actors of the space as small-scale and informal, based on personal relationships.

[1019] Interview with participant nr. 5 (2017, Essex).

[1020] Interview with participant nr. 7 (2017, Sheffield).

[1021] *Ibid*.

[1022] Interview with participant nr. 30 (2020, virtual).

B. Micro-Level Communication: Connections, Informal Communications and Small-Scale Interactions in the Research Data Sharing Space

Lower-ranked officials, i.e. analysts, play a central role in the research data sharing space because the regulatory issues of this space are often resolved through long-drawn communication, often informal and involving a small number of individuals, rather than by the diktat of a high-ranking official. Such communication and, often, bargaining are facilitated by pre-existing relationships between actors in the research data sharing space.[1023] After referring to the role of such relationships, I show that the crucial interactions which influence the implementation of data sharing frameworks and, thus, effectively shape the space happen at the 'micro' level, often outside of formal governance structures and policy headlines.

i. 'Knowing the right person': trust and relationships in the research data sharing space

The language used in data sharing frameworks is often technical and complex. Data custodians need to ensure that the provision of access conforms to demanding data protection principles, intermediaries need to deliver sophisticated training programs for researchers, and researchers themselves need to observe the strict rules of secure environments. Interviewees of this study, however, feel that that the implementation of data sharing frameworks is not so neutral and objective. Pre-existing relationships and the development of trust between participants in the space influence the way in which requirements are interpreted and issues are negotiated. As starkly put by an administrative data researcher with experience in the Scottish landscape:

> *'It's definitely the case that if you know the right person, you'll get things done faster.'*[1024]

They add that the Scottish Administrative Data Research Centre (intermediary body) was much more effective than its English counterpart in negotiating the establishment of research data

[1023] These dynamics are also reflected in the 'informal' historical development of the research data sharing space, see *supra* chapter 4, A.

[1024] Interview with participant nr. 29 (2020, phone).

sharing frameworks due to Scottish academics already having working relationships in place with Scottish data custodians:

> *'When it came to demonstrating success in terms of data access, we were probably already ahead of the game in that sense. We already had access to a lot of the social care data and the education data that was, really, kind of the prize.'*[1025]

Relationships between data custodians and researchers are often integral to triggering significant breakthroughs that allow a research data sharing procedure to materialise. The devolved countries, being smaller and allowing the development of closer relationships, are good examples in that regard. A policy-maker, who works for the SAIL Databank (custodian) in Wales, provides some background on the establishment of the databank's much-hailed framework, which largely relies on an algorithm that splits and de-identifies the used datasets.[1026] They recall that the 'connections' of some of the academics with people working for the NHS Wales Informatics Service (NWIS) (intermediary body) were crucial for NWIS to take the SAIL investment seriously and prioritise the development of the matching algorithm:

> *'I think (redacted) had some connections with them because they used to work on some projects in relation to clinical coding, you could easily check it with them. The people originally involved in the bid, they had some connections anyway.'*[1027]

A data analyst, who works for the Knowledge and Analytical Services (custodian body) of the Welsh government, claims that the possibility of data linkage, i.e. the linking of administrative datasets that belong to different data custodians,[1028] arose out of established communicative channels between civil servants and academics. They explain that some academics first started to explore data linkage techniques. Since Wales is a 'small country', civil servants who manage

[1025] *Ibid*.

[1026] SAIL Databank (n 612).

[1027] Interview with participant nr. 21 (2019, Swansea).

[1028] Harron et al (n 142).

datasets 'get to know the academics within their field quite well'.[1029] This makes it easier for Welsh academics to influence the scope of data sharing agreements and, more generally, the interpretation of data sharing frameworks: they 'know who to talk to' and 'have already got relationships established'.[1030]

Coming to Scotland, an administrative data researcher with experience in the Scottish landscape, explains that researchers have been largely relying on their relationships with individuals working for data custodians to influence the design of data sharing procedures. They invoke the example of one of their more experienced researcher colleagues who 'has been around' and 'has lots of experience and contacts'.[1031] To manage relationships, this colleague is keen on following up on 'institutional rumours about who is up and who is down' within custodian bodies to understand who they 'need to talk to' for data access purposes.[1032] Another researcher, with similar background, recalls the establishment of a research data sharing process for the first time by a relatively small local council in Scotland. They explain that this particular local council had pre-existing relationships with their research centre, due to some previous collaboration which involved statistical analysis. To highlight the significance of these relationships, the interviewee draws a contrast with another local council, which they had approached without a prior relationship for data access:

> *'We also approached another council, a large council. We approached them, had a meeting with them and they were very cold. They said, you know, it's a great idea, but we don't have people to do it and we didn't even chase that up.'*[1033]

Another researcher in the Scottish landscape recalls one of the, arguably few, examples where a smooth and timely research data sharing procedure was facilitated by the Department for Work and Pensions (DWP) (custodian). They attribute this to the previous relationship of one of their researcher colleagues with DWP officials:

[1029] Interview with participant nr. 25 (2019, Cardiff).

[1030] *Ibid*.

[1031] Interview with participant nr. 35 (2020, virtual).

[1032] *Ibid*.

[1033] Interview with participant nr. 37 (2020, virtual).

> '*I think it was an easier task for (redacted), because (redacted) was part of the research team that conducted the PSE survey for the DWP*.'[1034]

In Northern Ireland, like Wales, the success of the local Administrative Data Research Centre (ADRC) (intermediary) in agreeing the establishment of research data sharing procedures with data custodians is explained by virtue of the previous relationships between academics and civil servants. A data analyst, who works for the ADRC in Northern Ireland, recalls that their first administrative data research project came into fruition because their previous employer, a veterinary lab, had strong links with the government department that held some relevant data.[1035] An interviewee, who also works for the same ADRC in a policy-making capacity, explains how academics were often approached by civil servants who either knew them or heard about them and wanted to discuss the establishment of a data sharing process:

> *'We had people with data come to us, (...) and say because they would know that there are academic experts in their area (...) we have data, would you like to use it?'*[1036]

Although it is more difficult to establish trusting relationships in a bigger country like England, there are reasons to believe that pre-existing relationships have influenced the implementation of research data sharing frameworks in the English landscape. A policy-maker, who works for the Economic and Social Research Council (intermediary) and has a background as an administrative data researcher, recalls that someone from their research team had 'good personal relationships' with the department of Communities and Local Government that would often 'help unlock things'.[1037] A data analyst, who has worked for the Department for Education (custodian) and now works for an independent research organisation, explains that their personal relationships, from the time they worked for the department, are integral to them understanding how data sharing requirements will be interpreted:

[1034] Interview with participant nr. 38 (2020, virtual).

[1035] Interview with participant nr. 15 (2018, virtual).

[1036] Interview with participant nr. 10 (2018, Belfast).

[1037] Interview with participant nr. 19 (2018, Swindon).

> *'We have good contacts with DfE. Most people worked for me at one point or another and this comes as quite handy! Do I know how to do this with the HMRC? No. With the DWP? No.'*[1038]

In some cases, intermediary bodies may be better-placed than researchers in handling relationships with data custodians with a view to broadening research data sharing frameworks, consistent with their organisational mission to promote research in the public interest. A data analyst, who works for the SAIL Databank (custodian) in Wales, notes that the databank leverages its good relationship with the National Statistician for Wales to convince other Welsh data custodians to subscribe to its data sharing framework.[1039] Negotiations become easier when this key relationship is being invoked:

> *'We can overcome a lot of their risk aversion by telling them who else is confident that we are a safe pair of hands with the data (…) they were reassured by the fact that the Welsh government already had a long relationship with us.'*[1040]

If relationships are integral to progressing things in this space, the absence of relationships can hamper the development of a data sharing procedure. A policy-maker, who has worked for the Administrative Data Research Network (ADRN) (intermediary), attributes the data sharing difficulties faced by the ADRN in England to the lack of 'contacts' and 'good working relationships' with data providers.[1041] As their data analyst colleague working in Wales concurs, the more complicated organisational structures are, the more difficult relationship-building becomes:

> *'Certainly, you can't scale that model up to Whitehall very easily. DWP is eight times bigger itself than the Welsh government. There are many more factors in play, many more relationships.'*[1042]

[1038] Interview with participant nr. 30 (2020, virtual).

[1039] Interview with participant nr. 22 (2019, Swansea).

[1040] *Ibid*.

[1041] Interview with participant nr. 21 (2019, Swansea).

[1042] Interview with participant nr. 22 (2019, Swansea).

Relationships between data sharing actors are, then, critical to the development of research data sharing procedures. As it will now be shown, such relationships are, for the most part, not formed within formal organisational structures. Commonplace, informal interactions are the most prominent facilitators of influential communications within the research data sharing space.

ii. Informal interactions: shaping data sharing frameworks far from the spotlight

If connections and personal relationships are pivotal in developing research data sharing frameworks, there is a question as to *where* this process of development happens. As will be suggested here, the *locus* of the negotiating process is not to be found in formal channels of communication and governance structures, but in informal interactions.

Momentous policy and legal decisions, like the establishment of the ADRN in 2012,[1043] or the entry into force of the Digital Economy Act in 2017, are the first port of call for most writers reviewing the research data sharing landscape in the UK.[1044] Such decisions are, naturally, taken at the highest level of policy- or law-making and constitute the reflection of large organisational action over a long time. They reflect the collective intention of such large actors as the Department for Business, Enterprise and Skills or the UK Parliament to establish, amend or demarcate a framework governing research data sharing. Stepping for a moment, however, outside of the formal structures that have been set up for this purpose, one can notice that the dynamic development of data sharing framework often hinges on the commonplace and the informal.

A policy-maker, who has worked for the HM Revenue and Customs (HMRC) (custodian), claims that significant aspects of their department's uniform data sharing policy were decided in informal interactions with their colleagues:

[1043] Administrative Data Taskforce (n 25).

[1044] Laurie and Stevens (n 7); Jackson (n 315).

> *'Going out and talking to people and high-jacking data shares that were in progress or new requests for other areas and applying our governance (…) That brought consistency.'*[1045]

In designing the so-called 'front door' service at the HMRC, which allows broader data sharing for research purposes compared to the HMRC Datalab,[1046] the same interviewee admits that they were influenced by some 'early exploration of potential proposals with the ADRN' through informal meetings with academics.[1047] Another policy-maker, working for the Health and Social Care (custodian) department in Wales, welcomes the honesty and flexibility that is guaranteed by informal interactions, as opposed to formal and public meetings where everything is on the record. They feel that such interactions are the most productive in negotiating the development of a data sharing process with all interested stakeholders:

> *'In Wales we are small enough to put all these people in a room, often we have quite a good relationship in terms of engagement - it really is straightforward.'*[1048]

They add that, through these interactions, solutions can be given to thorny issues that would necessitate years of formal negotiations, had the official channels been used. For example, if a particular weakness in a research data sharing framework is to be ameliorated via legislative change, it would require 'a process of at least three years'.[1049] By informally liaising with the appropriate stakeholders, a 'consensus around consistent single interpretations' can be reached and 'cut through the complexity' that is inherent in a sophisticated, robustly governed activity like research data sharing.[1050]

A data analyst, working for the Knowledge and Analytical Services (custodian) of the Welsh government, recounts how significant their informal interactions with their higher-ups in the government were for the initial establishment of data linkage processes for research purposes.

[1045] Interview with participant nr. 32 (2020, Oxford).

[1046] Bega (n 431).

[1047] Interview with participant nr. 32 (2020, Oxford).

[1048] Interview with participant nr. 24 (2019, phone).

[1049] *Ibid*.

[1050] *Ibid*.

They describe the moment they started negotiating this with their line manager within the civil service:

> *'When I first went to our chief statistician at the time and said, we should do this, (redacted) was like: 'well, I've got no money'. I said to (redacted), can I go and get some money? Do you mind if I do that? And (redacted) basically said, literally, fill your boots. This is how I started talking to the ESRC.* '[1051]

Reflecting on that experience in the past, the same interviewee admits that they are still relying on informal interactions to move things forward and agree on the establishment of research data sharing procedures within the Welsh government. For example, some months ago, they had noticed that information governance officials within local authorities were reluctant to proceed with research data sharing due to worries about the limits imposed by the GDPR. To allay these worries, they brought these officers in touch with the Welsh government data protection officers, setting up a regular, weekly meeting to help 'building a common understanding of the GDPR and how they can share data for stats and research'.[1052]

An administrative data researcher, with experience in the Scottish landscape, elaborates on the many aspects of the research data sharing process that are predominantly shaped by informal interactions. To begin with, when constructing a research proposal that will satisfy custodian requirements as to the 'public interest', informal provision of advice from more experienced researcher to their younger colleagues is often crucial:

> *'A lot of that is quite informal, not because there isn't provision, but, inevitably, informal support is a crucial component of any organisation.* '[1053]

Following that stage, the extent to which the proposal will be deemed satisfactory by the decision-making panels within custodian bodies will, again, be significantly influenced by informal interactions. They recall a particular example where they managed to negotiate a data

[1051] Interview with participant nr. 25 (2019, Cardiff).

[1052] *Ibid*.

[1053] Interview with participant nr. 34 (2020, virtual).

sharing agreement thanks to their personal relationship with data analysts working for the custodian body holding the data:

> *'I mentioned the informal there and I think that actually was crucial in moving forward the work. Lots of things just happened to align (…) Through individuals within National Records of Scotland, we were really able to start moving forward, thinking how we would enable the data to be made available for linkage and later for analysis.'*[1054]

Based on this experience, they dismiss the view that 'more paperwork is the way forward'.[1055] An information governance official, working for the Department for Work and Pensions (custodian), seemingly shares the opinion of the previous interviewee. They stress that lots of clarifications and refinements are required when negotiating a data sharing process. Informal communication between the negotiating parties 'by telephone' allows all that to happen 'much more rapidly than it might happen if it was all written'.[1056] In a similar vein, a data analyst, who has worked for the Office for National Statistics (ONS) (custodian) in the past, recalls that the establishment of a data access procedure by the ONS happened after an informal, almost covert arrangement with one of their colleagues from IT:

> *'Luckily, I knew someone from ONS who was in IT and was very keen and was happy to work with me on this. We were lucky there, had no resistance at all. We basically pretended this was a pilot to get this up and running.'*[1057]

To say that informal interactions often shape the contents of data sharing frameworks is not to say that such interactions are always spontaneous or whimsical. Informality refers to the communications taking place out of the formal decision-making structures, not necessarily to the lack of organisation in such interactions. As a matter of fact, it may be the case that data sharing participants will set up informal teams, not belonging to a particular organisation or

[1054] *Ibid*.

[1055] *Ibid*.

[1056] Interview with participant nr. 8 (2017, Sheffield).

[1057] Interview with participant nr. 42 (2020, virtual).

employing professional staff, in the interest of giving some structure and coherence to the interactions.

A prime example of such an informal team is the so-called 'Administrative Data Forum' (ADF) in Northern Ireland. This is a *sui generis* body, in that it is comprised of representatives from data custodians, but it does not itself fall under any of the actors' categories of this space. An information governance official, who works for the Northern Irish Statistics and Research Agency (NISRA) (custodian), provides some background on the establishment and function of the ADF. As they explain, the forum is comprised of 'senior statisticians working for different government departments'[1058] and was established by NISRA with a view to stimulate discussion about research data sharing in the Northern Irish public sector.[1059] Crucially, data custodians were represented in the forum by their own data analysts. NISRA data analysts used this forum to 'build rapport' with their counterparts, thus securing critical 'buy-in' for the development of research-related access procedures.[1060]

Some interviewees reflect on the significance of the ADF in the establishment of research data sharing procedures in the Northern Irish public sector. A policy-maker, working for the Honest Broker Service (custodian) in Northern Ireland, recalls that they had participated in the past in the ADF. They welcomed the 'information sharing' capacity of the forum with regard to making better use of government data and producing policy-relevant research.[1061] A data analyst, working for the HBS, agrees that the forum has been 'very valuable' in making Northern Irish data custodians more comfortable with the idea of research data sharing:

> *'Once people come to grips with how it works in practice, a lot of the barriers will fall away I think. Having a forum where people can share information and address concerns is really important to this.'*[1062]

[1058] Interview with participant nr. 9 (2018, Belfast).

[1059] *Ibid;* NISRA were interested in doing so due to their mandate as the Northern Irish Statistics and Research Agency to promote statistical research.

[1060] *Ibid*.

[1061] Interview with participant nr. 13 (2018, Belfast).

[1062] Interview with participant nr. 16 (2018, phone).

Observing the utility of the ADF in establishing research data sharing frameworks in Northern Ireland, the Administrative Data Research Centre (ADRC) (intermediary) - Northern Ireland adopted the idea of establishing informal teams and groups to effectively manage aspects of the data sharing process. A data analyst, working for the ADRC-Northern Ireland, explains that they have agreed with NISRA analysts that they have a 'monthly team meeting (…) to discuss the state of play' for the progress of data access requests.[1063] A colleague of theirs adds that these meetings are not part of their official duties, yet they are always 'helpful' and 'well-structured', enabling them to stay on top of the most pressing data sharing challenges.[1064]

Similar informal groups operate in other countries. A data analyst, who works for the SAIL Databank and the ADRC in Wales, describes the operation of an informal group which aims to bring together administrative data researchers, policy-makers and data analysts from Welsh data custodians:

> *'We have developed the model (…) where one of my researchers, a policy person from Welsh government and a government researcher, they work as a team. The policy person has the emerging findings to make policy decisions, the government researcher is building the information base for the benefit of government research and the researcher here is deriving impact from the changes that the policy person is making.'*[1065]

As in Northern Ireland, then, Welsh and Scottish experience seems to reflect the significance of informal channels of communications like 'forums' as a facilitator of making data custodians more confident in research data sharing. As another interviewee, who works as a data analyst for the KAS of the Welsh government, explains, a body like the Northern Irish ADF exists in Wales. This body, called the Welsh Statistical Liaison Committee, 'brings together lots of the right professionals and lots of the local authorities',[1066] enabling closer cooperation and better chances of promoting the idea of research data sharing. An administrative data researcher, with experience in negotiating access with Scottish data custodians, describes the establishment of

[1063] Interview with participant nr. 15 (2018, virtual).

[1064] Interview with participant nr. 14 (2018, Belfast).

[1065] Interview with participant nr. 22 (2019, Swansea).

[1066] Interview with participant nr. 26 (2019, Cardiff).

an informal group to facilitate better communication between data custodians and researchers. Apart from them, there were data analysts and policy people 'with a particular interest in data sharing' from Police Scotland and the Fire Rescue Scotland.[1067] Although they concede that, in this particular case, the conversations were not fruitful and the attempt to establish a data linkage framework eventually 'lost its momentum',[1068] they express satisfaction for the efficiency of the communication outside of a long, formal process.[1069]

Informal interactions often unblock research data sharing negotiations and help participants overcome the state of standstill. As the next sub-section will show, interactions in the research data sharing space are also to a large extent routine and small-scale, with changes taking shape through long and incremental processes.

iii. Small-scale interactions: iteratively and incrementally shaping data sharing frameworks

Some interactions in the research data sharing space may receive significant attention and participation from interested stakeholders, being perceived as 'influential' and reflecting broader changes in the field.[1070] By and large, however, interactions in this space take place between a few individuals as part of a daily organisational routine and require numerous iterations before yielding a particular output for the regulation of data sharing decisions.

An interviewee, who has worked for the Office for National Statistics (ONS) (custodian) as a data analyst in the past, recalls a particular data linkage negotiation with officials working for the HM Revenue and Customs (HMRC) (custodian). They stress the significance of multiple iterations for 'allowing people to talk and express (…) and address their concerns'.[1071] In this particular case, they had to meet data analysts and policy-makers from the VAT and PAYE departments in the HMRC. Although initially happy, the custodians consulted with their colleagues and eventually came back with hesitations about the establishment of the research

[1067] Interview with participant nr. 35 (2020, virtual).

[1068] *Ibid*.

[1069] *Ibid*.

[1070] Thomas and Walport (n 2); Law Commission (n 2); Laurie and Stevens (n 7).

[1071] Interview with participant nr. 42 (2020, virtual).

data sharing process. The interviewee had to go at great lengths and engage in multiple conversations to reboot the negotiations:

> *'So, I said, ok, can I talk to your bosses? And I went through the tree, I had two audio conferences, two physical trips to London, basically until I got to the directors responsible for these departments within the HMRC.'*[1072]

As the interviewee recalls, it was not the substance of the conversations and their arguments that helped reassure data custodians and, eventually, convince them to proceed with the data sharing process. It was the *method* of engaging in conversations, iteratively refining the proposals and practically ensuring the custodians that all their worries are being carefully considered. In their own words:

> *'The conversations were exactly identical (...) Although there was no new information, I could have sent a document, but the document wouldn't have the same impact! (...) Doing all this in a personal way, particularly for senior people, is really, really important. Because you really want to know about all of this stuff. I'll have you sign this off - are you happy to do this?'*[1073]

The same interviewee recalls a series of iterations with the legal team at the ONS, when they were seeking to formalise the access of external researchers, who until then signed the so-called '1 pound contracts',[1074] to ONS data. Having some difficulties in the early conversations, they took some time to think and then 'went back' to the legal team at the ONS to frame their desired aim differently. This is when they realised that solutions to complex governance challenges may only come through multiple iterations:

> *'I went back to the lawyers and said, ok, can we treat academics as government researchers for the purposes of accessing this? And then they came up and said, yes, you do this, that and the other. That was the first time that I realised that*

[1072] *Ibid*.

[1073] *Ibid*.

[1074] *Ibid*, also see *supra* chapter 4 p 143 on the '1-pound contracts'.

> *getting the question right was really important and understanding how people are responding to the question.'*[1075]

This is a striking example of the importance of multiple iterations in the development of particular research data sharing frameworks. Even if the substance of the request was the same, i.e. allow academics as research users of the ONS data, the solution was only reached after numerous iterations that incrementally yielded a better mutual understanding. A policy-maker, who works for the Administrative Data Research Centre (ADRC) (intermediary body) in Scotland, describes a similarly iterative process of negotiating access to Police Scotland's data. They recall that one of the affiliated with the ADRC academics 'set up meetings with many people from Police Scotland (custodian) for a long time',[1076] aiming to make the custodians more comfortable with the ADRC's sponsored research. In four years' time, they remember 'at least 35 meetings',[1077] in which academic researchers would seek to demonstrate the impact that can be generated from establishing a research data access framework and the data custodians would express their concerns. Eventually, patience and persistence through these iterations paid off:

> *'This finally bore fruit in that now what Police Scotland have done is they realised that they need to have someone who actually has got the authority to release data across the board and (…) as we speak, they are setting up a couple of cautious projects to test the water.'*[1078]

Iterations are not only critical in overcoming barriers during data access negotiations. In some cases, they might be necessary to even *start* a data access negotiation. One reason might be that there is lack of researcher knowledge about the available datasets. As an administrative data researcher, with experience in the Scottish landscape, explains, it may be the case that multiple iterations are needed for a researcher to even understand what kind of data is available and could potentially be used by them. With many data custodians in Scotland not having 'any

[1075] *Ibid*.

[1076] Interview with participant nr. 31 (2020, virtual).

[1077] *Ibid*.

[1078] *Ibid*.

discoverability for their datasets',[1079] multiple inquiries to data analysts, who hold the datasets, become necessary to kickstart a data access request, especially when considering the level of detail that researchers have to include in their requests. Without this type of communication, or when such communication does not bear fruit, researchers are 'left in the dark'.[1080]

Small-scale iterations can also help to establish whether datasets are available in terms of custodians *desiring* to make them available for research use. An information governance official, who works for the electronic Data Research and Innovation Service (intermediary) in Scotland, notes that they are in 'regular contact' with data custodians, so that they can then inform researchers when 'something is available'.[1081] Another administrative data researcher with experience in Scotland adds that there is often a need to go through multiple layers of communication with numerous individuals to get a grasp of what would custodians actually make available:

> *'You need to send an email to one person, so that they can send an email to the second person and say they've been chased by us to show that there's some external pressure (…) I guess you have these organisational dynamics of people who are interested in different things within the Scottish government.'*[1082]

An additional reason for which iterations are necessary relates to the fact that there is lack of knowledge about the potential benefits that administrative data research can generate for a custodian body and, thus, insufficient incentives to establish a data sharing framework. A policy-maker, who has worked for the HM Revenue and Customs (HMRC) (custodian) in the past and currently works for the Cabinet Office, notes that they see repeated conversations with data custodians as the best way to help them realise how research data sharing can address policy gaps:

[1079] Interview with participant nr. 29 (2020, phone).

[1080] *Ibid*.

[1081] Interview with participant nr. 36 (2020, virtual).

[1082] Interview with participant nr. 35 (2020, virtual).

> *'Perhaps you are in a department and you have a fraud problem, but you might not necessarily know how to tackle it. We are trying to provide the little bit of extra information and handhold people through the governance process.'*[1083]

The same interviewee recalls a particular example where repeated iterations with MoJ officials enabled establishing a data sharing process which was at the risk of being aborted. In that case, the process would be originally established for the analytical benefit of a particular area within the Ministry of Justice (MoJ) (custodian) who at some point 'ran out of money to actually land it'.[1084] The interviewee recalls that they engaged in a series of conversations to 'salvage' this project and do not allow it to be perceived as a 'complete waste of money'.[1085] In their words:

> *'What I did was I went to the entire department, the MoJ, and asked them for the requirements of all their divisions. I knew that the prisoner probation service also had similar requirements - bringing them all together (…) we could finish the solution that another part of the MoJ had funded.'*[1086]

In this example, multiple iterations between small groups of individuals resulted in overcoming an impasse in a previously initiated research data sharing procedure. It may also be the case that iterations provide the opportunity of striking a compromise between the negotiating parties as to the conditions under which data will be shared. A policy-maker, working for the Department for Work and Pensions (DWP) (custodian), suggests that applicants that are patient and persistent enough to chase DWP analysts are more likely to eventually agree on a data sharing agreement. In those cases, researchers 'would never get a flat no coming out of the blue'; there would normally be some back-and-forth discussion on whether alternative variables can work for the researchers, so that DWP analysts will be able 'to take it further'.[1087]

A colleague of this interviewee, working in an information governance capacity for the DWP, agrees that agreeing on a data sharing agreement 'isn't just a yes or no, it is iterative,

[1083] Interview with participant nr. 32 (2020, Oxford).

[1084] *Ibid*.

[1085] *Ibid*.

[1086] *Ibid*.

[1087] Interview with participant nr. 7 (2017, Sheffield).

with revisions'.[1088] They stress that the 'human factor', through personal conversations, can often 'come in to get a way around' barriers to data sharing. Iterations between custodians and requesting parties are important to 'catchup' or 'retrofit' to avoid rejection of a request.[1089]

Even when iterations are internal within a custodian body, they can be of utmost importance for the eventual decision about a research data sharing procedure. A data analyst, who works for the Department for the Economy (custodian) in Northern Ireland, claims that 'scepticism' about research data sharing can 'certainly be overcome' through iterations between analysts and policy-makers.[1090] They explain that, upon reception of a data access request, they would often 'informally discuss' with their policy colleagues, asking them to 'raise any concerns' and try to address these concerns to move the negotiation forward.[1091] Another data analyst, working for the Office for National Statistics (custodian), confirms that the 'most important dialogue' is that between the Office's analysts and the Office's policy-makers who have the final word before an access request can be approved.[1092] They stress that it is of utmost importance that both sides have patience during this dialogue, engaging in a series of iterations in a spirit of 'co-operating and trying to work through a mutually reasonable decision at the end of the process'.[1093]

Apart from the more relational aspect of building rapport and allaying the worries of data custodians, repeated iterations are often necessary due to the complexities of research data sharing from a technical and information governance perspective. An information governance official, working for the NHS Wales Informatics Service (NWIS) (intermediary), explains that handling risks of reidentification from the use of administrative data necessitates regular disclosure controls and oversight of researcher practices. In some cases, researchers may receive initial approval for a particular use of the data, but 'follow-up' later to amend 'certain statistical considerations from the initial request'.[1094]

[1088] Interview with participant nr. 8 (2017, Sheffield).

[1089] *Ibid*.

[1090] Interview with participant nr. 16 (2018, phone).

[1091] *Ibid*.

[1092] Interview with participant nr. 2 (2017, Titchfield).

[1093] *Ibid*.

[1094] Interview with participant nr. 23 (2019, Cardiff).

> *'So, they would come back, but the purpose and the original methodology has changed. We need to consider this slightly differently now (...) you need to go back and reset some of this stuff (...) With Wales being a smaller place, we are able to have a closer conversation with people in that line.'*[1095]

Small-scale and repeated iterations, often as part of routine daily interactions between low-level officials working for data custodians, requesting and intermediary bodies, incrementally lead to the construction of research data sharing frameworks. Even if such frameworks then receive the formal legal form or the imprimatur of a high-level data sharing strategy designed by a large ministerial department, they are often shaped through these informal and iterative communicative channels.

C. Conclusion: The Importance of Commonplace, Low-Level Interactions

The 'rules of the game' in the research data sharing space are shaped through interactions that are more complicated than what meets the eye. In principle, high-ranked officials, like policy-makers, who work for data custodians are the ultimate decision-makers and dictate the direction in which research data sharing is headed, either resulting in broader and more permissive, or narrower and more restrictive frameworks.

A closer look, however, at the experiences of participants in the research data sharing process has illuminated the interactive and relational character of decision-making. Decisions are neither solely taken at the high-level, nor unilateral in nature. Low-ranked officials, the data analysts, play a significant role in framing the realities of decision-making and mediating the dialogue between decision-makers and requesters. Analysts are the experts, the mediators and the streamliners, often devising ingenuous solutions to enable higher-level officials to proceed with the creation and implementation of data sharing procedures.

Analysts can play such a pivotal role in unlocking the barriers to research data sharing despite their lack of formal decision-making powers because the interactions in the research data sharing space are generally not always pursued with a view to a high-level, landmark legal or policy decision. Long-term relationships and mutual trust between data sharing participants

[1095] *Ibid*.

offer critical assurances for a data sharing process to go ahead. Even if formally reflected in the legal framework or an officially published policy agenda, the requirements and the powers that constitute data sharing frameworks are constructed and negotiated in the commonplace, daily actions of the space's, often low-level, participants. Such interactions often happen outside of formal decision-making channels and may require numerous iterations before incrementally resulting in the rules that govern data access. The following, concluding chapter draws the tenets of the argument together and reflects on the future for the regulation of research data sharing decisions in the UK.

8. CONCLUSIONS AND RECOMMENDATIONS: THE PAST, PRESENT AND FUTURE REGULATION OF 'INDEPENDENT', DATA-DRIVEN RESEARCH

> *'I'm not saying it's going to happen tomorrow but (...) I'd say that if we in Government invest both the time and the money now, in perhaps another ten years we could have an analytical infrastructure that would be the envy of the world (...) I mean a range of administrative datasets flowing routinely from data management systems into secure settings (...), freeing up time for more detailed interpretation and analysis of the implications of research findings for our policy colleagues.'* [Lowe, 'The value of safe settings in evidence-based policy making'][1096]

After summarising the key findings of the present thesis about the past and present regulation of research data sharing in the UK, I provide, in this concluding chapter, a set of recommendations for the future of such regulation. These recommendations proceed from the normative premise that widening access to administrative data and enabling independent social-scientific research in the public interest should be pursued. This premise is required, as it has been shown that normative beliefs and assumptions about the desirability of independent public interest research have been shaping organisational decision-making and practices in this space. Proceeding from the particular normative premise is justified by virtue of the potential of independent public-interest research on administrative data to contribute to both greater transparency and efficiency of governmental policy-making and implementation. Finally, considering the limitations of the present study, I identify a set of main directions for future research in this area.

A. The Research Data Sharing Space: Concealing the Political Behind the Technical and the Commonplace

As it has been shown in this thesis, research data sharing in the UK has been and currently is regulated within a particular 'space', the boundaries of which have been (re)constructed and negotiated through the interaction of particular organisational actors, i.e. data custodians,

[1096] S Lowe, 'The value of safe settings in evidence-based policy making' (2018) <http://eprints.ncrm.ac.uk/4255/>.

intermediary bodies, research organisations and individual researchers making data access requests. If we read the arguments made in the previous chapters on the whole, certain cross-cutting conclusions about the regulation of research data sharing decisions in the UK public-sector are starting to emerge. Within this space, the 'political' is artfully concealed behind the 'technical' and the 'commonplace': well-established political views about e.g. public-sector efficiency and accountability,[1097] reflected in unwritten, conventional normative assumptions about the ownership of administrative data and the relationship of administrative data research with policy-making, determine the types of data sharing frameworks that are established and implemented. These normative assumptions, captured within a collective 'research data sharing culture', effectively shape the 'rules of the game' that filter access to administrative data for research purposes, being much more influential than legal rules and high-level policy decisions. Their formation often happens far from the policy-making spotlight: they are incrementally established and revised via small-scale interactions often between low-level technical staff, rather than momentous legal and policy decisions made by high-level officials or bodies.

Chapter 2 introduced the analytical framework adopted in this thesis, drawing on the concept of regulatory space. The present thesis has drawn on rich, original-primary and secondary mainly qualitative empirical data to provide an in-depth understanding of three main dimensions of the research data sharing space from the perspective of its participants: the *actors* which have populated the space and have been involved in shaping the data sharing frameworks over time, the *institutions* or 'rules of the game' that govern research access to public-sector administrative data, and the *interactions* through which the actors form their organisational identities by reference to the institutions, while at the same time creating these institutions themselves. Chapter 3 outlined the methodological framework of this thesis. Fourty-two (42) interviewees, working for a variety of organisations in the research data sharing space, have participated in this study, offering crucial insights and information that are not available in the public domain. Various documents, both from the public domain and disclosed by this study's

[1097] These views are political in the sense that they seek to maintain or re-balance a particular distribution of power, to crystallise and express particular organisational interests. This is the case either when a view is expressed by a data custodian body invoking new public management efficiency rhetoric to define legitimate research by reference to direct relevance to cutting down public spending, or when a research organisation invokes liberal ideals of holding State power accountable and scrutinised to bargain for wider data access.

interviewees, have also been used to critically examine and in part support the findings presented here.

The empirical chapters of this thesis unpack the three main dimensions of the research data sharing space. First, in chapter 4, the thesis provided a comprehensive analysis of the organisational *actors* that have participated in the historical development of the research data sharing space over the last four decades. From the early 1980s to the late 2010s, this analysis showed how particular organisations in the UK public sector, like the Office for National Statistics (ONS), the Department for Work and Pensions (DWP) and the Administrative Data Research Network (ADRN) sought to widen or narrow access to data for research purposes. Research data sharing, the analysis showed, started as an ad-hoc, informal activity, initially being led and conducted within the confines of public authorities. It gradually expanded to a closed circle of entrusted academics until the 2000s. With various actors in the data sharing space, such the ONS and the Cabinet Office, evangelising about the importance of widening access to independent social-science researchers for 'efficiency' and 'accountability' purposes, formal data sharing frameworks with concrete requirements and approval stages were established and their use became more widespread due to initiatives like the ADRN in the second half of the 2010s. However, it was shown, as we are moving forward, that research data sharing procedures remain fragmented across the UK, with major public-sector data custodians avoiding or actively resisting a publication of their criteria and requirements for granting access to researchers. This prompted critical questions about the 'rules of the game' that govern access to data for research purposes on the ground: to what extent do interpretations of legal powers and limits on sharing data determine the creation and application of transparent and predictable frameworks in practice? How do the interactions between the actors of this space, mainly government departments, academic or civil-society researchers and intermediary bodies, shape the actually applied requirements and processes for data access?

Second, in chapters 5 and 6, the thesis shifted its attention to the 'rules of the game' or *institutions* that govern research data access in the UK public sector. Seeking to interrogate claims in the data sharing literature, it proceeded from assessing the relevance of 'legal' institutions for the regulation of research data sharing decisions. A definition of 'legal' institutions, with the formal legal framework as its starting point, extended by the understandings of 'legal' powers and limits shared by the participants of this study, was adopted. Although lived experiences of research data sharing are not necessarily uniform in the

four UK countries, they significantly converged as to law's limited reach in determining the actual establishment and implementation of data sharing frameworks. In the different countries, participants may substitute the 'spirit of the law' with the 'spirit of the public benefit' (England), may pay lip service to the law's strictness (Northern Ireland), may use the law as a post-hoc justification for a particular data sharing procedure (Wales) or may highlight the communicative difficulties with lawyers representing data custodians (Scotland). In all cases, most of them would concur that legal interpretations are *not* the strongest factor shaping the process of research data sharing.

Having shown this, the analysis shifted, in chapter 6, to the most determinative factor behind the 'rules of the game' that govern data access, i.e. a collective 'research data sharing culture' that is formed by normative assumptions about the ownership of administrative data and the value of administrative data research that is relevant to policy-making. The multiple ways in which different 'research data sharing cultures' shape the creation and implementation of data sharing frameworks were discussed. In connection with data ownership, custodians who consider administrative data as a public good are more likely to establish structured data sharing procedures, becoming accountable to the research community and the broader public with regard to their data sharing criteria.[1098] Custodians who think of data as owned by them are more likely to resist accountability, thus avoiding to publicly commit themselves to particular data sharing criteria and processes. With regard to the relevance of research to policy-making, when custodians are satisfied that research satisfies their policy aims, they are more likely to set up a data sharing process, even committing themselves to specific decision-making steps. When, however, they are unclear as to the relevance or the compatibility of a proposed research request with their policy agendas, inertia, reluctance and caution are likely to emerge, hampering progress in the establishment or the transparent implementation of data sharing frameworks. In both cases, relevance to policy-making is likely to establish particular powers of custodians within the research data sharing frameworks, e.g. with regard to co-designing the research questions of the approved projects.

[1098] Such accountability is achieved through communication when it comes to researchers, as well as through publication of the data sharing procedure and relevant criteria on an organisation's website, see for example Office for National Statistics (n 150).

Third, in chapter 7, the discussion shifted to the last dimension of the research data sharing space, the *interactions* between the organisations that populate the space, through which the 'rules of the game' that filter access to administrative data are shaped. In theory, high-level policy makers are the ultimate decision makers who determine the contents of data sharing frameworks. In practice, however, it was shown that the formation of such frameworks takes place far from the spotlight and the high-level legal and policy decisions made by senior officials with formal decision-making competences. A closer look at the experiences of participants in the data sharing process illuminated the interactive and relational character of decisions. Decisions are neither solely regulated at the high-level by e.g. ministers or senior civil servants, nor are they unilateral in nature. Low-ranked officials, the data analysts, play a significant role in framing the realities of decision-making and mediating the dialogue between decision-makers and requesters. Analysts possess the expertise to develop solutions to support higher-level officials in creating and implementing data sharing procedures. Analysts can play such a pivotal role in unlocking the barriers to research data sharing despite their lack of formal decision-making powers because the interactions in the research data sharing space are generally not always pursued with a view to a high-level, landmark legal or policy decision. Even if formally reflected in the legal framework or an officially published policy agenda, the requirements and the powers that constitute data sharing frameworks are constructed and negotiated in the commonplace, daily actions of the space's, often low-level, participants. Such interactions often happen outside of formal decision-making channels and may require numerous iterations before incrementally resulting in the rules that govern data access.

The arguments made in this thesis aspire to engage constructively and move forward elements of both the nascent public-sector data sharing literature and the broader regulatory space literature. In respect of the literature on government data sharing for research purposes,[1099] the contributions relate to both breadth and depth. Considering that most of the so-far literature relies on high-level policy perspectives, often focused on central government intentions and experiences, the present thesis went deeper than identifying general-level potential influences (such as 'law' or 'culture') on research data access decisions. Instead, it offered an empirically-informed account of the manifestation of legal and cultural influences in a variety of organisations across the UK public sector, considering the developments in the

[1099] E.g. Law Commission (n 2); Thomas and Walport (n 2); Laurie and Stevens (n 7); Administrative Data Taskforce (n 25); Jackson (ns. 315 and 490); Elias (n 309).

regulation of such data sharing over the last four decades. This assessment has not only contributed to better evaluating the validity of existing literature views about the relative strength of legal or other obstacles to public interest research data sharing.[1100] It has also made political objectives, e.g. efficiency-driven new public management rationales or democratic ideals of government accountability, behind research data sharing regulation more visible. These findings shall, hopefully, both stimulate theoretical thinking in this area and inform practical considerations about such issues as initiating legal reform or using litigation strategies to achieve breakthroughs in research data access.[1101]

With regard to regulatory space analysis, this thesis has fleshed out the significance of low-level relationships and interactions, often between technical staff, as a significant analytical vehicle, potentially akin to such 'resources' as information, organisational capacity, wealth or legal authority,[1102] to explain developments in a regulatory space. Through that lens, the findings of this thesis highlight the potential centrality of unwritten and uncodified normative frames as the *locus* of regulatory conflict and bargaining. These contributions are presented by reference to the ways in which a collective 'research data sharing culture' shapes the formation of governance frameworks and, thus, the exercise of regulatory power in this space. In that sense, the present thesis departed from literature approaches which explain regulatory power struggles by reference to an actor's capacity to influence the interpretation of codified and formally articulated requirements.[1103] By capturing and examining subjects' perceptions of the influence of legal and cultural regulatory forces across the UK public-sector from the bottom-up, this thesis has also diverged from high-level approaches to analysing a regulatory space. In both cases, the aim has been to enrich the analytical toolkit of regulatory space studies.

[1100] Law Commission *ibid; cf* Laurie and Stevens *ibid*. The conclusions offered here have cast doubts upon conceptualising 'barriers' in a top-down way and seeking to overcome them through either legal or cultural shifts, even if culture shapes more saliently the 'rules of the game' that govern research data access compared to law.

[1101] *Infra* section B on reflections about the utility of litigation in the context of administrative data research. Also see similar findings about other similar regulatory spaces, e.g. Dove (n 12) 185-6 on the space of health research regulation and how high-level changes in the legal framework do not necessarily directly impact what is happening in the day-to-day practices of actors like research ethics committees.

[1102] Scott (n 47).

[1103] Jonnergård and Larsson-Olaison (n 49) 460; Canning and O'Dwyer (n 49) 169. See the discussion *supra* chapter 2, C.

The arguments supported by the findings of this thesis are not normative ones. They seek to provide in-depth insights and understanding of the way in which the conduct of research data sharing has been and currently is regulated in the UK, informed by the perspective of its own participants. Nonetheless, considering the great policy interest in regulating data processing and data ownership, it is helpful to briefly offer here some recommendations on the future of research data sharing regulation in the UK. These recommendations proceed from the premise that widening access to administrative data for independent social-science researchers is normatively desirable and serves the 'public interest'.

B. The Future for the Regulation of Research Data Sharing: Designing Legal and Cultural Shifts

What is the future of regulating access to administrative data for social-scientific research purposes in the UK public sector? The increasing digitalisation of civil service provision, including the potential automation of crucial administrative decisions, makes it likely that increased interest in the conditions for data access will be sustained. Both the government and civil society actors, as well as researchers, will likely grapple with the prerequisites for achieving a robust evidence-base in the four countries of the UK, with a view to allowing both the improvement of government policy-making and the independent scrutiny and evaluation of delivered policies.

First, a comprehensive strategy for making well-established cultural attitudes to research data sharing within the public sector more friendly to research is urgently needed. This is not to say that top-down declarations about e.g. data being a 'public good' would dramatically change the 'rules of the game' that govern data access on the ground. This work's findings have demonstrated that cultural attitudes to research data sharing are well-embedded and contingent on the organisational history and mission of a custodian body. Such attitudes are not easy to shift and would require long-term and concerted efforts from all stakeholders. Nonetheless, leadership should embrace the effort to shift the culture in the UK public sector, providing incentives for civil servants to become more transparent about their management of administrative data. Small steps can make bring substantial change, such as the requirement that research data sharing decision-making criteria and steps are published, as well as making publicly available comprehensive catalogues of the datasets that the departments hold and can be used by researchers. The perceptions that administrative data is a 'public good', to be

managed and shared in the name of the public, shall be gradually infused across the civil service.

What is the role of researchers and intermediary bodies in helping this incremental culture shift materialise? Relationships in the research data sharing space, although long-term and collaborative, are certainly the place where power is exercised. As it was previously noted in this thesis, both data custodians (e.g. the Department for Work and Pensions) and researchers or intermediary bodies (e.g. the Administrative Data Research Network) have been leveraging their resources to shift arrangements in their favour over time.[1104] With that in mind, there will certainly be merit in requesting parties seeking to leverage their resources, including the potential of litigation, to demand that research data sharing is democratised, i.e. that data custodians embrace more transparent and facilitative research data sharing procedures to enable wider access. There are, however, limits to what confrontational approaches to data access can yield. By contrast to investigative journalism, administrative data research is not interested in 'moments' but in sustained access to government data. Hostility can hamper the progressive development of a more research-friendly culture and bring resistance through the back door. Hence, until the cultural shifts bear their ripe fruits, requesting parties may want to consider adapting their negotiating approach by taking into account the prevalent normative attitudes with regard to research data sharing, as articulated in this study. Small steps can again bring the discussion forward, allowing valuable ideological cross-fertilisation between the different camps and bring the realisation of the research data sharing ideal into fruition in the long term.

Second, it is important to ask, can the law help move things forward? This thesis sought to reflect on whether we should reform the law to achieve the aims of widening access to administrative data for researchers. It was demonstrated that recommendations for legal reform in the research data sharing space, as a way of pursuing particular policy aims, need to be treated with caution. Governed by technically complex and long-drawn interactions among a variety of organisations, research data access is governed by norms whose interpretations are mediated through the framing of data analysts and are negotiated both within and outside of formal decision-making channels. This central finding cautions against attempting to prescribe particular obligations of data custodians to disclose data for research or justiciable rights of

[1104] *Supra* chapter 4.

researchers to require them to do so. Such legal rules, were they to be implemented, would face significant problems in application, with participants like data custodians using their interpretative latitude to resist disclosure when it is against their established views about e.g. their entitlements over administrative data or the appropriateness of research without a clear policy relevance as an activity that uses public resources.

That being said, it is also true that the current legal framework governing research data sharing in the UK is problematic. Administrative data ownership is not legally regulated,[1105] whereas sharing data for research is regulated by the broad and permissive discretionary powers established by the Digital Economy Act 2017. Although prescriptive legal regulation of the matter would face problems, it is important to provide more transparency and clarify the status of administrative data as a 'public good' in principle. Freedom of information (FOI) legislation can be an important locus of legal reform in that regard, as the current FOI Act 2000 is outdated for the purposes of facilitating access to administrative data for independent researchers. In previous times, FOI legislation enabled researchers to gain access to unpublished information about the design and implementation of government policies that affect individual citizens and groups. The current UK Freedom of Information (FOI) Act 2000, however, cannot adequately capture the nature of administrative data research which involves continuous information flows and requires on-going access to massive government datasets, which have to be regularly cleaned, curated and updated. The FOIA is not apt for achieving such sustained and comprehensive access, for reasons related both to its textual limitations and its scope. Reforming freedom of information legislation would be an important step towards facilitating independent research on digital experiments and reforms in the public sector.

C. Limitations and Main Directions for Future Research

Future interdisciplinary research in this area will be of utmost importance to provide further illumination of the way in which the topical activity of public-sector data sharing is regulated in the UK. As it has been mentioned in previous chapters, there are some limitations in the

[1105] Note that the terminology of 'ownership' is per se problematic, since administrative data are not legally the 'property' of data custodians according to a consistent line of UK Court of Appeal case law that shows that right-holders in respect of data do not have property in the data, see *supra* (n 321). Nonetheless, participants of the space are expressing views of 'ownership' and the analysis in this thesis has often reflected that.

present study, such as the limited amount of publicly available information about research data sharing and the difficulty to identify a stable 'population' of research data sharing participants.[1106] Considering that it is relatively early days for research data sharing in the UK, the population of research data sharing participants is amorphous, with many people changing positions after short periods and with public bodies being somewhat uncertain about the allocation of decision-making responsibilities with regard to data sharing. Especially with regard to the ADRN investment, which has been a focal point of reference for the present study, its shortcomings created a lot of nervousness which created additional barriers to accessing and recruiting interviewees. These limitations make further research all the more important.

As noted in this work's methodological chapter,[1107] the arguments made here neither seek to generalise about data sharing decisions in the whole of the UK public sector, nor claim to have captured all the nuances and differences that may characterise the decision-making process for some public bodies. By collecting comprehensive data about the way in which organisations have interacted in creating the research data sharing space, it has sought to lay the foundations for understanding the space, its inhabitants and the normative forces that regulate data access, without drawing sharp red lines to demarcate the space's boundaries in a definitive, nomological manner. Hopefully aided by the UK government's decision to provide more information about research data sharing decisions, future researchers shall be able to build on some of the insights presented here, refining or challenging them along the way.

There are productive lines of enquiry for future, interdisciplinary research into the regulation of research data sharing, in the UK and elsewhere. From a descriptive point of view, thicker descriptions of the data sharing process, potentially carried out by sociologists of organisations or political scientists, will provide much-needed clarity on the blurred lines between hierarchically superior and inferior data access decision-makers. Knowing more about the organisational environments will stimulate new lines of enquiry: what is it about specific public-sector organisations (e.g. ministerial departments) that consolidates a certain set of normative beliefs, to which succeeding generations of civil servants subscribe without radical differences? What is the role of leadership in instigating culture shifts from time to time? What is the impact of the organizational 'field' on specific organisations' structures and cultures?

[1106] *Supra* chapter 2, A.

[1107] *Supra* chapter 3.

From an explanatory point of view, drawing on growing knowledge about research data sharing, social scientists could assess the findings from this thesis, potentially coming from a different methodological standpoint. For example, more information about officials working for data custodians, as well as about the outcome of data access decisions, will allow the development of an explanatory research design that treats decision outcomes as its 'dependent' variable and the various candidate forces that are analysed in this study as its 'independent' variables. Such a design would employ mixed methodologies to evaluate the present account about the process of research data sharing decision-making. A first step towards developing mixed-methods in this field has been made in the present study through the design of a small-scale, pilot survey questionnaire.[1108]

Finally, yet importantly, from a normative point of view, researchers working at the intersections of law, ethics and new technologies will most likely have crucial questions to ask about the management of government administrative data in the decades to come. New technologies in the public sector are likely to rely on data-driven applications, including artificial intelligence, triggering critical questions about power imbalances engendered in unequal benefits from the uses of administrative data. Shall we, then, perceive such data as owned by public bodies or as a 'public good'? Does it make sense to think of the data as individual data, belonging to each and every citizen? Who should access the data and under which conditions? Should private companies be allowed to participate in research data sharing and, if yes, under which conditions? Shall governments rely on private datasets in the future, linking them to datasets gathered by public bodies? What is the role of law or other regulatory instruments in achieving those aims?

Regulating public-sector data ownership and management, as well as the entitlements of various parties over such ownership and management, is likely to be one of the major challenges for law and regulation in the 21st century. The research questions addressed by this thesis, as well as the ambitious research agendas that the future questions imply, will hopefully contribute to a better understanding of this challenge.

[1108] Appendices, E.

APPENDICES

A. Participant recruitment letter

An example of a message sent to a prospective interviewee:

Dear …

I hope you are very well.

My name is Stergios Aidinlis and I am currently finalizing my doctoral research in Socio-Legal Studies at the University of Oxford. You may recall a survey I had circulated in the ADR annual conference in Belfast, June 2018 about the dynamics of administrative data acquisition for ADRN purposes. In the last phase of my qualitative data collection, I am particularly interested in identifying the difficulties with developing consistent data sharing standards for (non-health) research data sharing across the public sector, as well as the implications of those difficulties for researchers. I found your recent article on the administrative data landscape in Scotland really illuminating and am now writing to ask whether you would be happy to consider discussing with me some aspects of your experience.

I would not expect the interview to last longer than 45-60 minutes. In case you are willing to consider this, I am very happy to provide more information about the work, my data security standards, institutional oversight, ethical review, etc. It goes without saying that no directly or indirectly identifiable data about my interviewees go into publications. The discussion is always abstracted to the level of the different frameworks and the different drivers behind data sharing. Many thanks in advance for your kind consideration,

Best regards,
Stergios

B. Participant information sheet



Perspectives on administrative data sharing for research in the United Kingdom

Participant Information Sheet

Researcher: Mr Stergios Aidinlis, Centre for Health, Law and Emerging Technologies (HeLEX),
Ewert House,
Ewert Place, Oxford, OX2 7SG

Tel: …
Email: stergios.aidinlis@law.ox.ac.uk

1. What is the purpose of this study?

To collect qualitative data on participants' perspectives and experience of administrative data sharing for research purposes.

2. Why have I been invited to take part?

You have been invited because of your previous engagement with research data sharing infrastructures in the UK public sector.

3. Do I have to take part?

No. It is entirely up to you whether you take part in this study or not.

4. What will happen if I do take part?

I will arrange a qualitative interview with you to discuss administrative data sharing for research. The interview will last between 45 and 60 minutes. There are no physical risks involved but you will be sharing your time and information.

Interviews will be organised at a location that suits you or by telephone or online face-to face communication tools depending on what is most desirable.

5. What happens if I change my mind?

You can end your participation at any point, including during the interview. You do not have to give a reason for your decision.

6. How will the information I provide be used?

A. All research data and records will be stored for a minimum retention period of 3 years after publication or public release of the work of the research.

B. The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research. If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives, to facilitate its use in future research. The thesis will be published open access.

C. In case you consent to that, the interview will also be audio recorded. The audio recording of the interview will be transcribed and a thematic analysis will be carried out. Findings will be written up and used in the researcher's thesis for his DPhil degree at the University of Oxford and any peer-reviewed publications arising from them. The researcher (S. Aidinlis) will be the data custodian during this time and will be responsible for ensuring the secure storage and confidentiality of the data.

7. What steps will be taken to protect my information?

Audio recordings will only be accessed by the researcher and will not be shared with anyone else without the prior explicit written permission of the participant. No identifiable data will be shared with third parties. Physical copies of the recordings (e.g. on SD cards) will be kept in a locked filing cabinet. Digital copies will be stored on a secure, password protected drive on the University of Oxford, Nuffield Department of Population Health server. The recordings will be transcribed by the researcher himself or an approved commercial transcription service with a non-disclosure agreement. Transcripts will be anonymised and assigned a code. A key linking the codes to the personal details of interviewees will be stored securely and kept separate from the transcripts.

Recordings and transcripts will be used for research purposes only. This will include using the analysis of this data, which can include selected quotes from the data set, in the researcher's DPhil thesis at the University of Oxford and any peer-reviewed publications arising from them. Quotes in

published works will be identified only by the code assigned to the transcript. It will be clear in any resulting publications that participants were involved in administrative data sharing for research.

Anonymised transcripts may be shared with other members of the HeLEX team for the purposes of verifying coding and analysis of the material.

8. Who has reviewed this study?

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. Ref No: R50823.

9. What if there is a problem?

If you have a concern about any aspect of this project, please speak to the researcher (S. Aidinlis, 075 35329006) who will do his best to answer your query. The researcher should acknowledge your concern within 10 working days and give you an indication of how he intends to deal with it. 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 (Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD). The chair will seek to resolve the matter in a reasonably expeditious manner.

Thank you for reading this sheet and considering whether you would be willing to participate.

C. List of participants

Participant Number	Role	Affiliation	Job Type
1	Data Custodian	ONS	Information Governance
2	Data Custodian	ONS	Data Analyst
3	Data Custodian	ONS	Data Analyst
4	Data Custodian	ONS	Policy-Maker
5	Intermediary Body	ADS / Uo Essex	Data Analyst
6	Intermediary Body	ADS	Policy-Maker
7	Data Custodian	DWP	Information Governance
8	Data Custodian	DWP	Policy-Maker
9	Data Custodian / Intermediary Body	NISRA	Information Governance
10	Intermediary Body	ADRC-NI / QUB	Data User
11	Data Custodian	Honest Broker Service (NI)	Information Governance
12	Data Custodian	Honest Broker Service (NI)	Policy-Maker
13	Data Custodian	Honest Broker Service (NI)	Policy-Maker
14	Intermediary Body / Researcher	ADRC-NI / Ulster University	Data Analyst / Data User
15	Intermediary Body / Researcher	ADRC-NI/QUB	Data Analyst / Data User

16	Data Custodian	Department for the Economy (NI)	Data Analyst
17	Intermediary Body	Wellcome Trust	Policy-Maker
18	Intermediary Body	Wellcome Trust	Policy-Maker
19	Intermediary Body	ESRC	Policy-Maker
20	Intermediary Body	ESRC	Policy-Maker
21	Data Custodian / Intermediary Body	SAIL	Policy-Maker / Data User
22	Data Custodian / Intermediary Body	SAIL	Data Analyst
23	Data Custodian / Intermediary Body	NWIS	Information Governance
24	Data Custodian	Health and Care Research Wales	Policy Maker
25	Data Custodian	KAS (Welsh Government)	Data Analyst / Data User
26	Data Custodian	KAS (Welsh Government)	Data Analyst / Data User
27	Intermediary Body	CAG	Information Governance
28	Data Custodian	ONS	Data Analyst
29	Researcher	Uo Edinburgh	Data User
30	Data Custodian / Researcher	DfE / NFER	Policy-Maker / Data User
31	Intermediary Body	ADRC-S / SCADR	Policy-Maker

32	Data Custodian	HMRC	Policy-Maker
33	Data Custodian / Researcher	DfE / EPI	Data Analyst / Data User
34	Intermediary Body / Researcher	SCADR / Edinburgh Napier University	Data Analyst / Data User
35	Researcher	Uo Edinburgh	Data User
36	Intermediary Body	eDRIS (Scotland)	Information Governance
37	Researcher	Edinburgh Napier University	Data User
38	Researcher	Uo Glasgow	Data User
39	Researcher	NHS Scotland	Data User
40	Researcher	FFT Education DataLab	Data User
41	Researcher	Uo Edinburgh	Data User
42	Data Custodian / Researcher	ONS / UWE Bristol	Policy-Maker / Data User

D. Interview guide

An example of an interview guide:

Part 1 | Background in Administrative Data Sharing for Research

i) Reflecting on your experience in the public and private sector, could you give me a brief overview of your contributions to developing and implementing data sharing frameworks for research purposes?

ii) Have you negotiated yourself access to government data for research purposes? If yes, could you tell me about this experience?

Part 2 | Developing Research Data Sharing Frameworks in the Public Sector

iii) Can you tell me about the development of the research data sharing policy of the ONS VML in 2004? E.g. reflecting on your personal contribution to it, the actors which were involved in it (e.g. analysts or policy-makers) and the main challenges throughout it.

iv) How would you assess the application of the 5 safes (or other relevant frameworks) to organisational realities in the UK public sector? Would you say the process is more or less formalised and predictable? How would you compare the UK with other countries in that regard?

v) From the perspective of a data provider, what do you see as the main incentives, as well as the main barriers, behind establishing a transparent research data sharing framework and consistently applying it in practice?

Part 3 | Managing Government – Academia Relationships

vi) How crucial would you say is the existence of a pre-existing relationship between particular data providers and the academia for the latter to develop a research data sharing policy?

vii) What would you identify as the key lessons from your experience in fostering links between academia and government for social science data-intensive research? E.g. reflecting on key requirements for success and still unresolved challenges.

viii) Have you had any involvement in major funding investments (such as the ADRN) which attempted to bring government departments and academic researchers closer to facilitate administrative data research? Do you have any reflections on the challenges faced by such investments or their future direction?

ix) How could academic researchers make the benefit from administrative data research seem less 'diffuse' while safeguarding the independence of their work?

Part 4 | Law and Organisational Culture

x) How central was the legal framework in your experience of developing or applying research data sharing frameworks in your different positions?

xi) Would you say that, overall, the law facilitates the aim of sharing more data for social research or rather complicates the whole process?

xii) Have you witnessed significant changes in the way government departments manage and share administrative data for research purposes? Have departments progressively become more 'open' and 'transparent'?

Part 5 | Additional

xiii) Are there any other aspects of administrative data sharing for research that you feel I have overlooked and that you want to comment on? E.g. reflecting on

developments in the near future, most desired changes in the field or initiatives that are needed most.

xiv) Are there any questions you would like to ask me?

E. Pilot survey

Questions on different factors that influence research data sharing decision-making:

i. Law

6. Would you agree or disagree with the following statements about the criteria that data providers take into consideration when deciding whether to grant a data access request?					
In the context of administrative data sharing for research…	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
Whether the data can be lawfully shared is a central concern for the data providers	❍	❍	❍	❍	❍
The legal framework regulating administrative data in the UK is more about imposing constraints on access to such data than facilitating access	❍	❍	❍	❍	❍
Data providers are not likely to even consider a researcher's request, unless they have been first convinced that there is a clear and explicit legal basis to share data	❍	❍	❍	❍	❍

ii. Policy relevance of research

6. Would you agree or disagree with the following statements about the criteria that data providers take into consideration when deciding whether to grant a data access request?					
In the context of administrative data sharing for research…	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
Data providers are more likely to grant a researcher's request, when the research proposal can demonstrate a clear connection with their own or another public body's policy making priorities	❍	❍	❍	❍	❍

iii. Perceptions of data ownership

7. Would you agree or disagree with the following statements about the reasoning of data providers?					
In the context of administrative data sharing for research…	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
Data providers are usually willing to provide the reasons behind their decision to reject a researcher's request	❍	❍	❍	❍	❍
Data providers are likely to seriously consider an appeal to their decision to reject a request by the researcher	❍	❍	❍	❍	❍
Data providers will often revise their initial decisions to reject a request in response to an appeal by the researcher	❍	❍	❍	❍	❍

iv. Trust in data sharing collaborators

8. Would you agree or disagree with the following statements about the relationships between data providers and requesting parties?					
In the context of administrative data sharing for research…	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
Submitting a Freedom of Information Act request is likely to lead to a subsequent refusal of a data access request	❍	❍	❍	❍	❍

Data providers are more likely to grant a data sharing request when approached by other civil servants compared to academics	❍	❍	❍	❍	❍
External researchers are likely to be asked to satisfy more demanding requirements compared to researchers who work for the data providers	❍	❍	❍	❍	❍
Individuals who have themselves a background as researchers are more likely to grant a researcher's request, when negotiating on behalf of a data provider	❍	❍	❍	❍	❍

Free text box:

In case the above categories are not applicable to your experience or you have additional comments to describe it, please write about it:

How many data sharing requests have	**N/a**	**None**	**1**	**2-3**	**4-5**	**6-7**	**8+**

you been involved in negotiating?							

Question to solicit interview participation:

9. Would you be willing to be interviewed about the issues raised by this survey by the researcher?
❍ Yes – please provide an e-mail address or a phone number for the researcher to contact you: ……. ❍ Not now

BIBLIOGRAPHY

Books and book chapters

Baldwin R, Cave M and Lodge M, *Understanding Regulation: Theory, Strategy and Practice* (OUP 1999)

Banakar R and Travers M (eds), *Theory and Method in Socio-Legal Research* (Hart 2005)

Bradley A W et al, *Constitutional and Administrative Law* (16th edn, Pearson 2015)

Bega D, 'The HMRC Datalab: Sharing administrative and survey data on taxation with the research community' in A Dusa et al (eds), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences* (Scivero 2014)

Braithwaite J and Drahos P, *Global Business Regulation* (CUP 2000)

Braun V and Clarke V, *Successful Qualitative Research* (Sage 2013)

Brint S and Karabel J, 'Institutional Origins and Transformations: The Case of American Community Colleges' in Powell and DiMaggio (eds), *The New Institutionalism in Organizational Analysis* (Chicago University Press 1991)

Bryman A, *Social Research Methods* (5th edn, 2016 OUP)

Burdon M et al, 'Access Control in Federated Databases: How Legal Issues Shape Security' in M Wimmer et al (eds), *Electronic Government: 6th International Conference, EGOV 2007* (Springer 2007)

Charmaz K, *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis* (Sage 2006)

Charmaz K and Bryant A, ‘Grounded theory and credibility’ in D Silverman (ed), *Qualitative Research* (3rd edn, Sage 2011)

de Vaus D, *Surveys in Social Research* (5th edn, Routledge 2001)

DiMaggio P, ‘Constructing an Organizational Field as a Professional Project: U.S. Art Museums, 1920-1940’ in Powell W and DiMaggio P (eds), *The New Institutionalism in Organizational Analysis* (Chicago University Press 1991)

Dove E, *Regulatory Stewardship of Health Research: Navigating Participant Protection and Research Promotion* (Edward Elgar 2020)

Edelman L, *Working Law: Courts, Corporations and Symbolic Civil Rights* (Chicago University Press 2016)

Erdos D, ‘Systematically Handicapped? Social Research in the Data Protection Framework’ (2011) 20 Information & Communications Technology Law 83

Ewick P and Silbey S, *The Common Place of Law: Stories from Everyday Life* (University of Chicago Press 1998)

Ferlie et al E, *The New Public Management in Action* (OUP 1996)

Fioretos O, Falleti T G and Sheingate A, ‘Historical Institutionalism in Political Science’ in theirs (eds), *The Oxford Handbook of Historical Institutionalism* (OUP 2016)

Fuller S, ‘The Social Construction of Knowledge’ in L McIntyre and A Rosenberg (eds), *The Routledge Companion to Philosophy of Social Science* (Routledge 2016)

Goerge R and Lee B J, ‘Matching and Cleaning Administrative Data’ in M Ver Ploeg et al (eds), *Studies of Welfare Populations: Data Collection and Research Issues* (NAP 2002)

Halliday S, *Judicial Review and Compliance with Administrative Law* (Hart 2004)

Hancher L and Moran M, ‘Organizing Regulatory Space’ in R Baldwin, C Scott and C Hood (eds), *A Reader on Regulation* (OUP 1998)

Hertogh M, *Nobody’s Law: Legal Consciousness and Legal Alienation in Everyday Life* (Palgrave 2018)

Hesse-Biber S and Leavy P, *The Practice of Qualitative Research* (2nd ed, Sage 2011)

Huntington S, *Political Order in Changing Societies* (Yale University Press 1988)

Jackson P, ‘From ‘’Intruders’ to ‘Partners’’: The Evolution of the Relationship Between the Research Community and Sources of Official Administrative Data’ in Crato N and Paruolo P, *Data-Driven Policy Impact Evaluation* (Springer 2019)

Jay R, *Data Protection Law and Practice* (Sweet & Maxwell 2012)

Keyser T and Dainty C, *The Information Governance Toolkit: Data Protection, Caldicott, Confidentiality* (CRC 2016)

King N and Horrocks C, *Interviews in Qualitative Research* (2nd edn, Sage 2019)

Lange B and Haines F, ‘Regulatory Transformations: An Introduction’ in Lange B, Haines F and Thomas R (eds), *Regulatory Transformations: Rethinking Economy-Society Interactions* (Hart 2015)

Lincoln Y and Guba E, *Naturalistic Inquiry* (Sage 1985)

Lynskey O, *Foundations of EU data protection law* (OUP 2015)

Mason J, *Qualitative Researching* (Sage 1996)

McLaughlin K et al (eds), *New Public Management: Current Trends and Future Prospects* (Routledge 2005)

Meyer J and Scott W, *Organizational Environments: Ritual and Rationality* (Sage 1983)

Miles M, Huberman M and Saldana J, *Qualitative Data Analysis: A Methods Sourcebook* (Sage 2013)

Moisander J and Valtonen A, *Qualitative Marketing Research: A Cultural Approach* (Sage 2006)

Morgan B and Yeung K, *An Introduction to Law and Regulation* (CUP 2012)

Powell W and Colyvas J, 'Microfoundations of Institutional Theory' in Greenwood R et al, *The Sage Handbook of Organizational Institutionalism* (Sage 2008)

Ritchie F, 'Access to business microdata in the UK: dealing with the irreducible risk' in Eurostat, *Monographs of official statistics: Work session on statistical data confidentiality* (2005)

Ritchie J and Spencer L, 'Qualitative data analysis for applied policy research' in A Huberman and M Miles (eds), The qualitative researcher's companion (Sage 2002)

Scott R, *Institutions and Organizations* (Sage 1995)

Silverman D, *Interpreting Qualitative Data* (Sage 2015)

Skeggs B, 'Situating the Production of Feminist Ethnography' in M Maynard and J Purvis (eds), *Researching Women's Lives from a Feminist Perspective* (Taylor & Francis 1994)

Thelen K and Steinmo S, 'Historical Institutionalism in Comparative Politics' in Thelen K, Steinmo S and Longstreth F (eds), *Structuring Politics: Historical Institutionalism in Comparative Analysis* (CUP 1992)

Tolbert P and Zucker L, 'What are Microfoundations? Why and how to study them?' in P Haack, J Sieweke and L Wessel, *Microfoundations of Institutions* (Emerald 2020)

Tomlinson J, *Justice in the Digital State: Assessing the Next Revolution in Administrative Justice* (Bristol Policy Press 2019)

Vibert F, *The New Regulatory Space: Reframing Democratic Governance* (Edward Elgar 2014)

Wade W and Forsyth C, Administrative Law (OUP 2014)
Weiss R, Learning from Strangers: The Art and Methods of Qualitative Interview Studies (The Free Press 1994)

Woolard M, 'Administrative Data: Problems and Benefits. A Perspective from the United Kingdom' in A Dusa et al (eds), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences* (Scivero 2014)

Journal articles

Aberbach J and Rockman B, 'Conducting and Coding Elite Interviews' (2002) 35(4) PS: Political Science & Politics 673

Almunia M et al, 'Expanding access to administrative data: the case of tax authorities in Finland and the UK' (2019) 26 Int Tax Public Finance 661

Arzberger P et al, 'Promoting Access to Public Research Data for Scientific, Economic and Social Development' (2004) 3 Data Science Journal 135

Bell J et al, 'Balancing Data Subjects' Rights and Public Interest Research: Examining the Interplay between UK law, EU human rights law and the GDPR' (2019) 1 European Data Protection Law Review 43

Black J, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (2001) 54 Current Legal Problems 103, 109

Black J, 'New institutionalism and naturalism in socio-legal analysis: Institutionalist approaches to regulatory decision making' (1997) 19(1) Law & Policy 51

Black J, ‘Regulatory Conversations’ (2002) 29 Journal of Law and Society 163

Braun V and Clarke V, ‘Using Thematic Analysis in Psychology’ (2006) 3 Qualitative Research in Psychology 77

Brett C and Deary I, ‘Realising health data linkage from a researcher’s perspective: Following up the 6-Day Sample of the Scottish Mental Survey 1947’ (2014) 5(3) Longitudinal Life Course Studies 283

Brower R, Abolafia M and Carr J, ‘On Improving Qualitative Methods in Public Administration Research’ (2000) 32(4) Administration & Society 363

Campos J et al, ‘The Challenges of Cybersecurity Frameworks to Protect Data Required for the Development of Advanced Maintenance’ (2016) 47 Procedia CIRP 222

Canning M and O’Dwyer B, ‘The Dynamics of a Regulatory Space Realignment: Strategic Responses in a Local Context’ (2013) 38 Account Org Soc 169

Carter P et al, ‘The Social License for Research: Why Care.data Ran into Trouble’ (2015) 41 Journal of Medical Ethics <https://jme.bmj.com/content/41/5/404>

Cascade E et al, Combining Electronic Health Records and Patient-Reported Information in the UK: Preliminary Results from the WASPS (Wales SAIL + PRO Study’ (2012) Value in Health

Chowdhury G et al, ‘Measuring the Impact of Research: Lessons from the UK’s Research Excellence Framework 2014’ (2016) 11(6) PLoS ONE.

Chowdury N and Wessel R, ‘Conceptualizing Multilevel Regulation in the EU: A Legal Translation of Multilevel Governance?’ (2012) 18(3) European Law Journal 335

Connelly R et al, ‘The role of administrative data in the big data revolution in social science research’ (2016) 59 Social Science Research 1

Cooper D, ‘Local government legal consciousness in the shadow of juridification’ (1995) 22(4) Journal of Law and Society 506

Corbin J and Strauss A, ‘Grounded Theory Research: Procedures, Canons, and Evaluative Criteria’ (1990) 13(1) Qualitative Sociology 3

Cowan D, ‘Legal Consciousness: Some Observations’ (2004) 67 Modern Law Review 928

Dyke S O et al, ‘Registered access: authorizing data access’ (2018) European Journal of Human Genetics (published online)

Einav L and Levin J, ‘The Data Revolution and Economic Analysis’ (2014) 14(1) Innovation Policy and the Economy 1, 7- 8

Ewick P and Silbey S, ‘Narrating social structure: stories of resistance to legal authority’ (2003) 108(6) American Journal of Sociology 1334

Ford D et al, ‘The SAIL Databank: building a national architecture for e-health research and evaluation’ (2009) 9 BMC Health Services Research

Hall P and Taylor R, ‘Political Science and the Three New Institutionalisms’ (1996) Political Studies 936

Halliday S, ‘After Hegemony? The Varieties of Legal Consciousness Research’ (2019) 28(6) Social & Legal Studies 859

Hand D J, ‘Statistical challenges of administrative and transaction data’ (2018) 181(3) Journal of the Royal Statistical Society Series A (Statistics in Society) 1

Harron K et al, ‘Challenges in administrative data linkage for research’ (2017) 4(2) Big Data & Society

Janghorban R, Latifnejad Roudsari R and Taghipour A, 'Skype interviewing: The new generation of online synchronous interview in qualitative research' (2014) 9(1) International Journal of Qualitative Studies on Health and Well-being

Jones K H et al, 'The other side of the coin: Harm due to the non-use of health-related data' (2017) 97 International Journal of Medical Informatics 43

Jones K H et al, 'The Good, the Bad, the Clunky: Addressing Challenges in Using Administrative Data for Research' (2019) 4(1) International Journal of Data Population Science

Jonnergård K and Larsson-Olaison U, 'Developing Codes of Conduct: Regulatory Conversations as Means for Detecting Institutional Change' (2007) 29(4) Law & Policy 460

Kaye J and Gibbons S, 'Mapping The Regulatory Space For Genetic Databases And Biobanks In England And Wales' (2008) 9(2) Medical Law International 111

Lane J, 'Building an Infrastructure to Support the Use of Government Administrative Data for Program Performance and Social Science Research' (2017) 1 The ANNALS of the American Academy of Political and Social Science 240

Lange B, 'Compliance Construction in the Context of Environmental Regulation' (1999) 8(4) Social & Legal Studies 549

Lange B and Cook C, 'Mapping a Developing Governance Space: Managing Drought in the UK' (2015) 68 Current Legal Problems 229

Laurie G et al, 'Charting Regulatory Stewardship in Health Research: Making the Invisible Visible' (2018) 27 Cambridge Quarterly of Healthcare Ethics 333

Laurie G and Stevens L, 'Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom' (2016) 43 Journal of Law & Society 360

Lodge M, ‘Regulation, the Regulatory State and European Politics’ (2008) 31(1) West European Politics 280

Lo Iacono V, Symonds P and Brown D, ‘Skype as a Tool for Qualitative Research Interviews’ (2016) 21(2) Sociological Research Online 103.

Lyon F et al, ‘Opening access to administrative data for evaluating public services: The case of the Justice Data Lab’ (2015) 21(2) Evaluation 232

MacKenzie R and Martinez Lucio M, ‘The colonisation of employment regulation and industrial relations? Dynamics and developments over five decades of change’ (2014) 55(2) Labour History 189

Malsch B and Gendron Y, ‘Reining in Auditors: On the dynamics of power surrounding an ‘’innovation’’ in the regulatory space’ (2011) 36 Accounting, Organizations and Society 456

March J and Olsen J, ‘The New Institutionalism: Organizational Factors in Political Life’ (1984) 78(3) The American Political Science Review 734

McBarnet D and Whelan C, ‘The Elusive Spirit of the Law: Formalism and the Struggle for Legal Control’ (1991) 54(6) Modern Law Review 848

McGrail K and Jones K, ‘Population Data Science: The science of data about people’ (2018) 3(4) International Journal of Data Population Science

Merry S E, ‘Legal Pluralism’ (1988) 22 Law and Society Review 869

Meyer J and Rowan B, ‘Institutionalized organizations: formal structure as myth and ceremony’ (1977) 83 American Journal of Sociology 340

Morris H et al, ‘Challenges of administrative data linkages: experiences of Administrative Data Research Centre for England (ADRC-E) researchers’ (2018) 3(2) International Journal of Population Data Science 97

Mourby M et al, ‘’Are ‘pseudonymised’ data always personal data? Implications of the GDPR for administrative data research in the UK’’ (2018) 34(2) Computer Law & Security Review 22

O’Reilly D et al, ‘Administrative Data Research Northern Ireland (ADR NI)’ (2020) International Journal of Data Population Science

Ospina S, Esteve M and Lee S, ‘Assessing Qualitative Studies in Public Administration Research’ (2017) 78(4) Public Administration Review 593

Perry A, ‘The Crown’s administrative powers’ (2015) Law Quarterly Review 652

Powell W and Dimaggio P, ‘The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields’ (1983) 48(2) American Sociological Review 147

Ramos E M et al, ’A mechanism for controlled access to GWAS data: experience of the GAIN Data Access Committee’ (2013) 92 American Journal of Human Genetics 479

Ritchie F, ‘Business Data Linking – Recent UK experience’ (2004) 33(1-2) Austrian Journal of Statistics 89

Ritchie F, ‘Secure access to confidential microdata: four years of the Virtual Microdata Laboratory’ (2008) 2 Economic & Labour Market Review 29

Ritchie F, ‘UK release practices for official microdata’ (2010) 26(3) Statistical Journal of the IAOS 103

Rumbold J and Pierscionek B, ‘The Effect of the General Data Protection Regulation on Medical Research’ (2017) 19(2) Journal of Medical Internet Research 47

Scott C, ‘Analysing Regulatory Space: Fragmented Resources and Institutional Design’ (2001) Public Law 283

Seitz S, ‘Pixilated partnerships, overcoming obstacles in qualitative interviews via Skype: a research note’ (2016) 16(2) Qualitative Research 229-235

Sexton A et al, ‘A balance of trust in the use of government administrative data’ (2017) 17(4) Archival Science 305

Silbey S, ‘Studying Legal Consciousness: Building Institutional Theory from Micro Data’ (2018) 3 Droit et société 698

Stevens L, ‘The Proposed Data Protection Regulation and its Potential Impact on Social Sciences Research in the UK’ (2015) 2 European Data Protection Law Review 97

Suddaby R, ‘How communication institutionalizes: A response to Lammers’ (2011) 25(1) Management Communication Quarterly 183

Taylor M J et al, ‘Developing a data sharing framework: a case study’ (2014) 8 Transforming Government: People, Process and Policy 151

Taylor M J, ‘Legal Bases for Disclosing Confidential Patient Information for Public Health: Distinguishing between Health Protection and Health Improvement’ (2015) 23(3) Medical Law Review 348

Teng J et al, ‘Sharing linked data sets for research: results from a deliberative public engagement event in British Columbia, Canada’ (2019) 4(1) International Journal of Data Population

Thatcher M and Coen D, ‘Reshaping European regulatory space: An evolutionary analysis’ (2008) 31 West European Politics 806

van der Wal Z, ‘Mandarins versus Machiavellians? On Differences between Work Motivations of Administrative and Political Elites’ (2013) 73(5) Public Administration Review 749

Waithira N et al, ‘Data management and sharing policy: the first step towards promoting data sharing’ (2019) 17 BMC Medicine

White R and van Koten H, ‘Co-designing for Sustainability: Strategizing Community Carbon Emission Reduction Through Socio-ecological Innovation’ (2016) 19(1) The Design Journal 25

Williamson O, ‘Comparative Economic Organization: The Analysis of Discrete Structural Alternatives’ (1991) 36 Administrative Science Quarterly 269

Wood S et al, ‘The Interactive Dynamics of Transnational Business Governance: A Challenge for Transnational Legal Theory’ (2015) 6(2) Transnational Legal Theory 333

Young J, ‘Outlining regulatory space: Agenda issues and the FASB’ (1994) 19(1) Accounting, Organizations and Society 83

Zucker L, ‘Institutional Theories of Organization’ (1987) 13 Annual Review of Sociology 443-464

Policy reports

Administrative Data Taskforce, ‘The UK Administrative Data Research Network: Improving Access for Research and Policy’ (December 2012)

Byrom N, ‘Digital Justice: HMCTS data strategy and delivering access to justice Report and recommendations’ (October 2019) <https://research.thelegaleducationfoundation.org/blog/digital-justice-hmcts-data-strategy-and-delivering-access-to-justice>

Cabinet Office, ‘Open Data: unleashing the potential’ (28 June 2012)
Law Commission, `Data Sharing Between Public Bodies: A Scoping Report' (2014) <http://www.lawcom.gov.uk/app/uploads/2015/03/lc351_data-sharing.pdf>

Centre for Market and Public Organisation, ‘NPD User Group’ <http://www.bristol.ac.uk/cmpo/npd-user-group/>

Expert Advisory Group on Data Access, ‘Governance of Data Access’ (June 2015) <https://wellcome.ac.uk/sites/default/files/governance-of-data-access-eagda-jun15.pdf>

HM Government, ‘Open Data White Paper: Unleashing the Potential’ (June 2012)

National Institute for Social Care and Health Research Wales, ‘Maximising the Use of Routine Data for Research in Wales’ (2013) <https://www.healthandcareresearch.gov.wales/uploads/Policy%20%26%20Strategy/Use%20of%20routine%20data/Maximising%20the%20Use%20of%20Routine%20Data%20for%20Research%20in%20Wales.pdf>

OECD, ‘The Path to Becoming a Data-Driven Public Sector’ <https://www.oecd-ilibrary.org/governance/the-path-to-becoming-a-data-driven-public-sector_059814a7-en>

REFORM, ‘Making NHS Data work for Everyone’ <https://reform.uk/research/making-nhs-data-work-everyone>

REFORM, ‘Sharing the Benefits: How to Use Data Effectively in The Public Sector’ <https://reform.uk/research/sharing-benefits-how-use-data-effectively-public-sector>

Shakespeare S, ‘Shakespeare Review: An Independent Review of Public Sector Information’ (May 2013)

Thomas R and Walport M, `Data Sharing Review Report' (2008) <https://webarchive.nationalarchives.gov.uk/+/http:/www.justice.gov.uk/docs/data-sharing-review.pdf>

UN Special Rapporteur on Extreme Poverty and Human Rights, ‘Digital technology, social protection and human rights: Report’ (2019) <https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx>

Wellcome Trust, ‘Data Policy Models for Funding Bodies’ (September 2018) <https://wellcome.ac.uk/sites/default/files/data-policy-models-for-funding-bodies-eagda.pdf>

Websites and miscellanea

ADRC Scotland, ‘From a Culture of Caution to a Culture of Confidence: Lessons Learned From Implementing the Public Records (Scotland) Act 2011’ <https://www.research.ed.ac.uk/portal/files/27327583/Publicity_Events_ADRC_Workshop_FINAL.pdf>

ADRN, ‘History’ <https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/adrn/>

ADR UK, ‘Administrative Data Research UK’ <https://www.adruk.org/>

ADR UK, ‘Administrative data is an invaluable resource for public good. Let’s use it’ (2019) <https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADRUK_Annual_Report_2018-19.pdf>

Aidinlis S et al, ‘Building a Justice Data Infrastructure Opportunities and Constraints’ (2020) (AI FOR ENGLISH LAW UKRI GRANT) <https://www.law.ox.ac.uk/sites/files/oxlaw/ukri_justice_data_report_fv_0.pdf>

Allsopp C, ‘Review of Statistics for Economic Policymaking: Final Report to the Chancellor of the Exchequer, the Governor of the Bank of England and the National Statistician’ (2003)

Article 29 Working Party, ‘Guidelines on consent under Regulation 2016/679’ (28 November 2017) <ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030>

Ball J and Marland M, ‘Male Earnings Mobility in the Lifetime Labour Market Database’ (1996) (Analytical Services Division, Department of Social Security Working Paper Series)

Besse P and Laurent B, 'From Statistician to Data Scientist' (2015) arXiv:1509.07804 [stat.OT]

Birmingham Safeguarding Children Board, 'Serious Case Review in respect of the death of Keanu Williams' (2013)
<https://www.cscb-new.co.uk/downloads/Serious%20Case%20Reviews%20%20exec.%20summaries/SCR_Archive/Birmingham%20SCR%20-%20KW%20(2013).pdf>

Booth R, 'Benefits system automation could plunge claimants deeper into poverty' (The Guardian) 14 October 2019 <https://www.theguardian.com/technology/2019/oct/14/fears-rise-in-benefits-system-automation-could-plunge-claimants-deeper-into-poverty>

Cabinet Office, 'Departments, agencies and public bodies' <https://www.gov.uk/government/organisations>

Centre of Excellence for Information Sharing, 'Information Governance' <https://informationsharing.org.uk/infogovernance/>

Chatsiou K et al, *Data Sources for Administrative Justice Research: A Preliminary Scoping Report* (2017) <https://doi.org/10.6084/m9.figshare.5303317.v2>

Deetjen U, *Internet use and health: a mixed methods analysis using spatial microsimulation and interviews* (DPhil thesis, University of Oxford 2016)

den Braak V et al, 'Trusted third parties for secure and privacy-preserving data integration and sharing in the public sector' (2012) Proceedings of the 13th Annual International Conference on Digital Government Research 135

Department for Culture, Media & Sport, 'Digital Economy Bill – Factsheet – Digital Government: Research and Statistics (clauses 56-68)'
<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535021/11._Research_and_Statistics.pdf>

Department for Digital, Culture, Media & Sport, 'UK National Action Plan for Open Government 2019-2021' (2019)

Department for Education, 'DfE external data shares' (18 July 2019) <https://gov.uk/government/publications/dfe-external-data-shares>

Department for Work and Pensions, 'DWP Open Data: Case studies' (2013) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/221159/dwp-open-data-case-studies.pdf>

Department for Work and Pensions, 'Research at DWP' <https://www.gov.uk/government/organisations/department-for-work-pensions/about/research>

Department for Work and Pensions, 'Work and Pensions Longitudinal Study' <https://www.gov.uk/government/statistics/work-and-pensions-longitudinal-study>

Department of Health, 'Research Governance Framework for Health and Social Care' (2nd edn, 2005)

Desai T, Ritchie F and Welpton R, 'Five Safes: designing data access for research' (UWE Economics Working Paper Series) (2016)

Early Career Researchers Using Scottish Administrative Data, 'eCRUSADERS: About' <https://ecrusad.co.uk/>

Edelman L, 'Legal Endogeneity and the Limits of Equal Opportunity' (2004) <https://memphis.edu/benhooks/creative-works/pdfs/edelman.pdf>

Education Policy Institute, 'About Us' <https://epi.org.uk/our-mission-and-purpose>

Endicott T, 'Parliament and the Prerogative: From the Case of Proclamations to Miller' (2016) Judicial Power Project <http://judicialpowerproject.org.uk/wp-content/uploads/2016/12/Endicott-2016-Parliament-and-the-Prerogative.pdf>

ESRC, 'Administrative Data Research Network' <https://esrc.ukri.org/research/our-research/administrative-data-research-network/>

ESRC, 'Big Data Network' <https://esrc.ukri.org/research/our-research/big-data-network/>
European Data Protection Board, 'Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (art. 70.1.b))' (23 January 2019) <edpb.europa.eu/sites/edpb/files/files/file1/edpb_opinionctrq_a_final_en.pdf>

Flanagan O, 'Opening up admin data to research: an interview with Melanie Wright' <https://www.statslife.org.uk/features/1689-opening-up-administrative-data-to-research-an-interview-with-melanie-wright>

Flynn M, 'South Gloucestershire Safeguarding Adults Board, Winterbourne View Hospital, A Serious Case Review' (2012) https://hosted.southglos.gov.uk/wv/report.pdf

Gill C, *Administrative justice and the control of bureaucratic decision-making* (PhD thesis, University of Glasgow 2016)

Government Statistical Service, 'Introducing the Administrative Data Methods Research Programme' (1 July 2019) <https://gss.civilservice.gov.uk/news/introducing-the-administrative-data-methods-research-programme/>.

Hales J et al, 'Evaluation of Employment Zones: Report on a Cohort Study of Long-Term Unemployed people in the Zones and a Matched Set of Comparison Areas' (2003) (National Centre for Social Research

Hand D, 'The Role of the ADRN Governing Board' (2013) <https://esrc.ukri.org/files/research/administrative-data-taskforce-adt/the-role-of-the-adrn-governing-board-david-hand/>

Health Research Authority, 'Confidentiality Advisory Group' <https://www.hra.nhs.uk/approvals-amendments/what-approvals-do-i-need/confidentiality-advisory-group/>

Health and Social Care NI, 'Honest Broker Service' <http://www.hscbusiness.hscni.net/services/2454.htm>

HMCTS, 'Access to courts and tribunals for academic researchers' <https://www.gov.uk/guidance/access-to-courts-and-tribunals-for-academic-researchers>

HMCTS Chief Executive, 'Modernising the Courts and Tribunals Service: Future of Justice Conference' (14 May 2018) <https://www.ucl. ac.uk/laws/sites/laws/files/ucl_foj_01_03_acland-hood.pdf>

HM Government, 'Open Government' <https://www.gov.uk/government/collections/open-government>

HM Government, 'How to access Department for Education (DfE) data extracts' (2018) <https://www.gov.uk/guidance/how-to-access-department-for-education-dfe-data-extracts>

HM Revenue and Customs, 'Research at HMRC' <https://www.gov.uk/government/organisations/hm-revenue-customs/about/research>

House of Commons, 'The Bichard Inquiry Report' (2004) <https://dera.ioe.ac.uk/6394/1/report.pdf>

House of Commons Science and Technology Committee, 'The big data dilemma: Fourth Report of Session 2015-16' <https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf>

House of Lords Constitution Committee, 'Constitution – First Report' (2000-01) <https://publications.parliament.uk/pa/ld200102/ldselect/ldconst/11/1101.htm>

Jackson P, 'The Route to Better Knowledge for a Better Society' (2017) <https://prezi.com/6pe-8m2u9ad0/the-route-to-better-knowledge-for-a-better-society/?utm_campaign=share&utm_medium=copy>

Jones P and Elias P, 'Administrative data as a research resource: A selected audit' (Report to the ESRC) (2006)

ICO, 'Anonymisation: managing data protection risk code of practice' <https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf>

ICO, 'Consent' <ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/consent>

ICO, 'Data Sharing Checklists' <https://ico.org.uk/media/for-organisations/documents/1067/data_sharing_checklists.pdf>

ICO, 'Data Sharing Code of Practice' <https://ico.org.uk/media/fororganisations/documents/1068/data_sharing_code_of_practice.pdf>

ICO, 'Principle (e): Storage limitation' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-limitation/>

Iveson M and Deary I, 'Navigating the landscape of administrative data in Scotland' (2019) 4 Wellcome Open Research 97

Kilkelly U, 'The right to respect for private and family life' (COE 2003)

Laurie G and Sethi N, 'Information Governance Of Use Of Health-Related Data In Medical Research In Scotland: Current Practices And Future Scenarios' (SHIP Core Programme 2 Working Paper No. 1)

Lewis P et al, 'Cambridge Analytica academic's work upset university colleagues' (2018) <theguardian.com/education/2018/mar/24/cambridge-analytica-academics-work-upset-university-colleagues>

Lindner C, 'Where do Data Scientists Come From?' <https://medium.com/indeed-engineering/where-do-data-scientists-come-from-fc526023ace>

Littlechild J, 'Developments in access to DWP data' (2015) <https://esrc.ukri.org/files/research/administrative-data-taskforce-adt/developments-in-access-to-dwp-data-joanna-littlechild/>

Lowe S, 'The value of safe settings in evidence-based policy making' (2018) <http://eprints.ncrm.ac.uk/4255/>

Manzoni J, 'Big data in government: the challenges and opportunities' (21 February 2017) <https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities>

Ministry of Justice, 'Evaluating our reforms: Response to PAC Recommendation 4, January 2019' <https:// www.gov.uk/government/news/moj-response-to-public-accounts-committee-transforming-courts-and-tribunals>

National Archives, 'Information Management Assessment: Department for Work and Pensions' (2015) <https://www.nationalarchives.gov.uk/documents/information-management/dwp-ima-report-2015.pdf>

National Audit Office, 'Challenges in using data across government' (21 June 2019) <https://nao.org.uk/wp-content/uploads/2019/06/Challenges-in-using-data-across-government.pdf>

National Records of Scotland, 'Welcome to Scotland's Census' <https://www.scotlandscensus.gov.uk/>

Office for National Statistics, 'ONS research and data access policy'

<https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy>

ONS, 'Data Science Campus' <https://www.ons.gov.uk/aboutus/whatwedo/datasciencecampus>

Open Democracy, 'Under pressure, UK government releases NHS COVID data deals with big tech' <https://www.opendemocracy.net/en/ournhs/under-pressure-uk-government-releases-nhs-covid-data-deals-big-tech/>

Pattaro S et al, 'Labour Market and Wider Impacts of Benefit Sanctions: A Scoping Review' (2019) 17th Annual ESPAnet Conference, Stockholm, Sweden, 5-7 Sept 2019

Population Data Science, 'Pioneering Population Data Science' <https://popdatasci.swan.ac.uk/about-us/>

Research Project Management, 'Administrative Data Research Network (ADRN) Mid-Term Review Report' (8 November 2016) <https://esrc.ukri.org/files/research/research-and-impact-evaluation/adrn-mid-term-review-report-8-november-2016/>

Richards N, 'Introducing the Administrative Data Research Centre Wales (ADRC-W)' (Blogs Wales Institute of Social & Economic Research, Data & Methods)

Ritchie F and Welpton R, 'Sharing risks, sharing benefits: Data as a public good' (Conference Paper) (2011 Eurostat)

Ritchie F and Welpton R, 'Addressing the human factor in data access: incentive compatibility, legitimacy and cost-effectiveness in public data resources' (UWE Economics Working Paper Series) (2014)

Robinson G et al, 'Public Attitudes to Data Sharing in Northern Ireland: Findings from the 2015 Northern Ireland Life and Times survey' (2018) (ADRC-Northern Ireland)

SAIL Databank, 'The Anonymisation Process' <saildatabank.com/saildata/data-privacy-security/#anonymisation-process>

Scottish Centre for Administrative Data Research, 'Data for social good' <https://www.scadr.ac.uk/administrative-data/data-social-good>

Scottish Government, 'Linking social care, housing & health data' (2010) <https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2019/01/health-community-care-data-linkage/documents/linking-social-care-housing-and-health-data-social-care-clients-and-patients-views/linking-social-care-housing-and-health-data-social-care-clients-and-patients-views/govscot%3Adocument/Linking%2BSocial%2BCare%252C%2BHousing%2Band%2BHealth%2BData%2BSocial%2BCare%2BClients%2527%2Band%2BPatients%2527%2BViews.pdf>

Scottish Government, 'Linking social care, housing support & health data for statistics: Project Plan and Privacy Impact Assessment (PIA)' <https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2019/01/health-community-care-data-linkage/documents/linking-social-care-housing-support--health-data---privacy-impact-assessment/linking-social-care-housing-support--health-data---privacy-impact-assessment/govscot%3Adocument/Linking%2Bsocial%2Bcare%252C%2Bhousing%2Bsupport%2B%2526%2Bhealth%2Bdata%2B-%2BPrivacy%2BImpact%2BAssessment.pdf>

Scottish Government, 'Data linkage for research in Scotland' <https://www.gov.scot/publications/data-linkage-for-research-in-scotland/pages/scottish-government-funded-linkage-projects/>

Scottish Longitudinal Study Development & Support Unit, 'Introduction to the Scottish Longitudinal Study' <https://sls.lscs.ac.uk/about/>

Skinner C, 'Statistical Disclosure Control for Survey' (2009) <http://personal.lse.ac.uk/skinnecj/SDC%20for%20survey%20data%20S3RI.pdf>

Smith G et al, 'The Value of Linked Administrative Records for Longitudinal Analysis' (ESRC National Longitudinal Strategy Committee) (2004)

UK Administrative Justice Institute, 'About UKAJI' <https://ukaji.org/about/>

UK Statistics Authority, 'First Meeting of the Administrative Data Research Network Board' (2014) <https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-minutesandpapersfromadrnboardon14april201_tcm97-43879.pdf>

UK Statistics Authority, 'Meet the ADRN Board' <https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/adrn/meet-the-adrn-board/>

UK Statistics Authority, 'Thirteenth Meeting of the ADRN Board: Agenda and Papers' (2017) <https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/06/ADRN-05-06-2017.pdf>

UK Statistics Authority, 'Research Code of Practice and Accreditation Criteria' (1 March 2018)

Understanding Patient Data, 'Putting people at the centre of decisions about patient data' <https://understandingpatientdata.org.uk/>

Wellcome Trust, 'Policy and Advocacy team' <https://wellcome.ac.uk/about-us/teams/policy-and-advocacy-team>

Welpton R and Wright M, 'HMRC Datalab: Engaging with the External Research Community' (2011) (ESRC)

Welsh Government, 'Fuel poverty data linking project: the impact on health of the Welsh Government Warm Homes Schemes' <https://gov.wales/fuel-poverty-data-linking-project-impact-health-welsh-government-warm-homes>

White M, ‘Macro evaluation of the New Deal for Young People’ (2000) (PSI briefing note).

Zamenopoulos T and Alexiou K, ‘Co-design As Collaborative Research’ (2018) (Connected Communities Foundation Series, Bristol) <http://oro.open.ac.uk/58301/1/Co-Design_CCFoundationSeries_PUBLISHED.pdf>