Sharing infectious disease related data and biological samples from low resource settings: perspectives on ethics and governance

A thesis submitted for the degree of Doctor of Philosophy

By Ruchi Baxi



Nuffield Department of Population Health

Wolfson College

University of Oxford

Trinity Term 2020

Acknowledgements

I am deeply grateful to so many people who have supported me to complete this thesis.

Supervisors
Thank you to my supervisors, Professor Mike Parker, Associate Professor Patricia Kingori and Dr Premila Webster for your encouragement and support over the course of this DPhil – I have learnt so much. Mike and Patricia, I am grateful to you for the time you have spent enthusiastically engaging with this project, discussing ideas and reviewing drafts. Premila, thank you for your encouragement to explore new avenues, that led me to the world of bioethics.

Collaborators
Thank you to the Infectious Diseases Data Observatory, especially Professor Philippe Guerin, Laura Merson, and Alice Hawryszkiewycz. Your support was essential to this project. My thanks also to all of the interviewees who agreed to participate in this study, for generously sharing your insights and experiences with me.

Academic support
I am grateful for the encouragement and advice from all of those at the Ethox Centre. It is such a friendly and stimulating place to work.

My thanks to those who took time to comment on drafts of the thesis: Stella Botchway; Frances Butcher; Stephanie Johnson; Mira Schneiders; and Halina Suwalowska. A big thank you to my sister Sweta, and my husband Henry for doing the painstaking task of proof-reading the thesis.

Thank you to Nia Roberts at the Bodleian Health Care Libraries for your invaluable support in selecting databases and developing the search strategy for the systematic scoping review.

I would also like to thank Tony Hope for our discussions on the thesis, and life in general. I always come away seeing things in a new light.

Funding
I would like to thank the Nuffield Department of Population Health for funding the fieldwork for this project.

Public Health Teaching Team
A big thank you to Associate Professor Adrian Smith, and Alison Cheshire for your understanding and encouragement over the course of this project. I especially appreciate your flexibility in trying different working patterns as the project moved through different phases.

Family and friends
I am so grateful to have such a supportive family and set of friends, I really could not have done this without you all. Your support is crucial at the best of times, and even more so in these times of uncertainty during this pandemic.

Henry, I am so thankful for your kindness, humour and patience, you always know what I need before I do. Thanks also to my daughter Leela for your boundless energy and joie de vivre, you bring me so much joy and perspective. I am so grateful to my parents, and the wider Baxi, Lee and Simms families for your unwavering practical and emotional support, and tireless enthusiasm for my work. And finally, a huge thanks to all of my friends for being there, listening to my worries and making me laugh.

Contents

ABSTRACT **8**

1. INTRODUCTION **11**

1.1. Infectious diseases as a global threat **11**

1.2. Data and sample sharing, and bioethics **12**

1.3. My personal motivation to pursue this thesis **13**

1.4. Thesis overview **14**

2. LITERATURE AND POLICY REVIEW **16**

2.1. Background **16**

2.2. Literature review: Methods **16**
2.2.1. Search methods: Peer-reviewed literature 17
2.2.2. Search methods: Policy documents 18
2.2.3. Data Collection and Analysis 19
2.2.4. Limitations of the review methods 21

2.3. Scoping review: Findings **22**
2.3.1. What does sharing data and samples mean in the infectious disease context? 22
2.3.1.1. Who is sharing *22*
2.3.1.2. What is shared? *23*
2.3.1.3. How does sharing take place? *23*
2.3.2. Reasons given to share data and samples 24
2.3.3. Concerns raised with sharing data and samples 25
2.3.3.1. The individual patient or research participant level *25*
2.3.3.2. The professional level *25*
2.3.3.3. The global level *26*
2.3.4. Ethical principles and considerations for practice 27
2.3.4.1. Ethical principles *27*
2.3.4.2. Consent *28*
2.3.4.3. Privacy *29*
2.3.4.4. Ownership *30*
2.3.4.5. Benefit sharing *32*
2.3.4.6. Trust *34*
2.3.4.7. Governance of data and sample sharing *34*
2.3.5. The policy context 36
2.3.5.1. The International Health Regulations 2005 ..... *37*
2.3.5.2. Global Health Security *38*
2.3.5.3. Convention on Biological Diversity and the Nagoya Protocol *39*
2.3.5.4. The Pandemic Influenza Preparedness Framework *40*
2.3.5.5. WHO Consultation: Developing global norms for sharing data and results during public health emergencies *42*
2.3.5.6. WHO ethics guidance *42*
2.3.5.7. Joint statement on data sharing in public health emergencies *45*
2.3.5.8. Médecins Sans Frontières Data Sharing Policy *45*
2.3.5.9. The International Severe Acute Respiratory and Emerging Infection Consortium Sample and Data Sharing Policy *47*
2.3.5.10. Key points from the policy review *48*

2.4. Conclusion 49
2.4.1. Scoping review conclusions 49
2.4.2. Research gaps 50

3. METHODS 52

3.1. Aim and objectives 52

3.2. Choice of research method 53
3.2.1. Possible choices of method 53
3.2.2. Rationale for using qualitative methods in this study 54

3.3. Research context 56
3.3.1. The Infectious Diseases Data Observatory – background 56
3.3.2. Visceral Leishmaniasis data sharing platform 57
3.3.3. Ebola Virus Disease data sharing platform 58

3.4. Data collection 59
3.4.1. Data sources – an overview 59
3.4.2. Semi-structured interviews 59
3.4.2.1. Sampling Frame 59
3.4.2.2. The process of recruitment 61
3.4.2.3. Conducting interviews 61
3.4.2.4. Reflections on face-to-face and remote interviews 65
3.4.2.5. Completion of interview data collection 66
3.4.3. Meeting observation 67

3.5. Data analysis and presentation of findings 68
3.5.1. Approach to analysis 68
3.5.2. Presentation of analysis in the thesis 72

3.6. Ethical considerations and clearance 72

3.7. Reflexivity: Reflections on my role as a researcher 73

3.8. Conclusion 76

4. TYPOLOGY OF DATA AND SAMPLES, AND THE ETHICAL IMPLICATIONS OF SHARING THEM 77

4.1. Typology: Data and sample collection 77
4.1.1. Differences between data and samples 78
4.1.1.1. Samples as a form of data 78
4.1.1.2. Samples distinct from data 78
4.1.1.3. Data and samples: Differing perceptions 79
4.1.2. Types of data and samples 80
4.1.2.1. Different types of samples 81
4.1.2.2. Research and non-research data and samples 83
4.1.2.3. Qualitative and quantitative data 83
4.1.2.4. Genetic/genomic data and samples 84
4.1.2.5. Aggregate and individual level data 85
4.1.2.6. Anonymised, de-identified and identifiable individual level data and samples 87

4.2. Sharing of data and samples 87
4.2.1. Data and sample sharing with known others 88
4.2.2. Data and sample sharing with potentially unknown others 90
4.2.2.1. Open access data 92
4.2.2.2. Managed access repository 95
4.2.3. Sharing different types of data and samples 97
4.2.3.1. Sharing samples compared to sharing data 97
4.2.3.2. Research and non-research data and samples 99
4.2.3.3. Qualitative data 100
4.2.3.4. Aggregate data 101
4.2.3.5. Anonymised, de-identified and identifiable data and samples 102

4.3. The infectious disease context 104
4.3.1. Outbreak and non-outbreak situations 105
4.3.2. Social stigma of disease 108
4.3.3. Diseases with associations with poverty 109

4.4. Conclusion 111

5. ORGANISATIONS AND PROFESSIONS, AND THEIR DIFFERING PRIORITIES 113

5.1. Role of organisations and professions 113
5.1.1. The World Health Organization 114
5.1.1.1. The WHO as a leader and source of guidance 114
5.1.1.2. The WHO as a locus of knowledge sharing 115
5.1.2. Organisations directly involved in sharing 116
5.1.2.1. Governments and public health professionals 116
5.1.2.2. Academic organisations and researchers 117
5.1.2.3. Health care providers, including non-governmental organisations 118
5.1.2.4. Commercial organisations 120
5.1.3. Organisations indirectly involved in sharing 121
5.1.3.1. Funders 122
5.1.3.2. Journals 124

5.2. Organisational and professional priorities, and emerging tensions 127
5.2.1. Public health priorities 128
5.2.1.1. Saving lives and improving health 128
5.2.1.2. Sharing early and rapidly 129
5.2.2. Academic priorities 130
5.2.2.1. Academic recognition 130
5.2.2.2. Maximising use of funding 131
5.2.2.3. Avoiding wasted data and samples 133
5.2.2.4. Contested priorities of academia and researchers 134
5.2.3. Health care and academic organisations: Protecting patients and research participants 135
5.2.4. Reputation and economy of the country 136
5.2.5. Commercial priorities 137

5.3. Priorities and low- and middle-income country settings 138

5.4. Conclusion 141

6. MEETING RESPONSIBILITIES AND MAINTAINING CONTROL 144

6.1. Responsibilities of those involved in sharing 144
6.1.1. Responsibilities to people from whom data and samples were collected 145
6.1.1.1. Duty of care: recognising the people behind the data and samples 145
6.1.1.2. Responsibilities to maintain privacy and confidentiality 148
6.1.1.3. Responsibilities to share benefits 151
6.1.2. Responsibility to the wider population 153
6.1.3. Responsibilities to other professionals involved in sharing 155

6.2. Maintaining control 156
6.2.1. Manifestations of control 156
6.2.1.1. Ownership as a form of control 156
6.2.1.2. Gatekeeping of access to data and samples as a form of control 158
6.2.2. Reasons for seeking control 160
6.2.2.1. Control to meet responsibilities 160
6.2.2.2. Control to maintain ties with the data and samples 163
6.2.2.3. Seeking control to avoid negative assessment of data quality by others 166

6.3. Conclusions 167

7. ENABLING ETHICAL DATA AND SAMPLE SHARING: THE ROLE AND LIMITS OF CONSENT, COMMUNITY ENGAGEMENT, TRUST AND GOVERNANCE 169

7.1. Actions prior to sharing 169
7.1.1. Seeking consent 170
7.1.1.1. The importance of consent 170
7.1.1.2. Different forms of consent 171
7.1.1.3. Consent – the importance of context 172
7.1.2. Community Engagement 177
7.1.2.1. The role of community engagement 177
7.1.2.2. Concerns with community engagement 178
7.1.3. Research ethics committees 180
7.1.3.1. The role of research ethics committees 180
7.1.3.2. Concerns about research ethics committees 182

7.2. Trust and confidence between professionals at the point of sharing 185
7.2.1. The importance of trust 186
7.2.2. Trust and location 188
7.2.3. Gaps left by trust 189

7.3. Governance for sharing and future uses 191
7.3.1. The role of governance for sharing and future uses 191
7.3.2. Agreements between sharing parties 192
7.3.2.1. The role of agreements 192
7.3.2.2. Concerns with agreements 195
7.3.3. Access Committees 198
7.3.3.1. The importance of access committees 198
7.3.3.2. Concerns with access committees 201

7.4. Conclusions 201

8. DISCUSSION 204

8.1. Introduction 204

8.2. Examination of thesis findings 206
8.2.1. Delineating the objects, the way they are shared, and the disease 207
8.2.1.1. Ethical significance of the types of data and samples 207
8.2.1.2. The ethical significance of the difference between sharing data and samples with known and unknown people 210
8.2.1.3. The ethical significance of different disease contexts 211
8.2.1.4. Thinking about the object, the way they are shared, and the disease 213
8.2.2 Organisations and professions, and their differing priorities 214
8.2.2.1. The range of organisations, and their roles 214
8.2.2.2. Differing priorities of different organisations and professions 215
8.2.2.3. The power of funders and journals 218
8.2.3. Ethics at the individual level: A sense of personal responsibility 220
8.2.3.1. Responsibilities to the wider population 220
8.2.3.2. Ethical tensions between meeting responsibilities to the people data and samples have been collected from, and those who might benefit from their use 221
8.2.3.3. Ownership and control, and meeting responsibilities 222
8.2.4. Enabling ethical data and sample sharing 223
8.2.4.1. Consent and community engagement 223
8.2.4.2. Committees and agreements 225

8.3 Perspectives on what constitutes 'ethical practice' during infectious disease responses 226
8.3.1. Well-recognised ethical considerations in data and sample sharing 227
8.3.2. Factors influencing ethically permissibility of sharing 227
8.3.3. Considerations that may not be framed as 'ethical' 228
8.3.4. Multiple perspectives on what constitutes ethical practice 230

8.4. Recognising divergent perspectives: the role of 'boundary objects' 232

8.5. Implications for policy and practice 237

8.6. Strengths and limitations of this study 239

8.7. Future research 241

8.8. Conclusions 242

REFERENCES 244

APPENDICES 256

Appendix 1: Key words and MeSH terms 256

Appendix 2: Medline search strategy 258

Appendix 3: Mind map of policy documents 260

Appendix 4: Participant information sheet 261

Appendix 5: Consent Form 263

Appendix 6: Interview guide 264

Appendix 7: List of codes 266

Abstract

The sharing of data and samples are increasingly promoted as crucial to the timely response to infectious disease outbreaks by bodies such as the World Health Organization. Data and sample sharing are often discussed as an ethical imperative that should occur in order to better understand diseases, and facilitate the development of timely innovations to respond to infectious diseases. However, data and sample sharing present a number of macro- and micro-level ethical considerations which need to be reconciled in practice. These include issues of consent, privacy, and the distribution of benefits. Furthermore, the urgent need to share and respond to infectious disease emergencies may also strain certain areas of best practice. Crucially, the perspectives of stakeholders involved in sharing data and samples in infectious disease contexts have not been well documented.

This study aims to address the lack of emphasis on stakeholders' views by foregrounding their perspectives and practical experiences of ethical issues and views on the role of ethics governance. To achieve these aims, I employed a qualitative methodology, including participant observation and semi-structured interviews to capture the experiences of 28 key stakeholders who were purposively sampled to represent a range of different roles and positions. The primary focus of this research was on gaining insights about sharing data and samples from low- and middle-income countries, given the disproportionate burden of infectious diseases in these settings.

The findings, and subsequent analysis, illustrate that data and sample sharing in the infectious disease context take place within a complex moral world. The importance of sharing data, and the optimal use of samples, was underscored by interviewees in this study. However, the nature of the data, samples, disease and the modes of sharing were all reported to have particular ethical implications. A further finding is that sharing often takes place between a diverse set of individuals and organisations, who can be unfamiliar with one another and who at times have differing

priorities, and make divergent arguments about how sharing ought to take place for it to be considered ethical. Outbreaks were noted to present particular challenges. Timeliness of data and sample sharing was considered ethically important in outbreak situations, but there were otherwise conflicting views on whether and how the urgent nature of the response ought to change practice. The pressures of an outbreak response also present practical limitations to meeting the usual requirements of good practice, such as obtaining informed consent.

The findings show that aspects of practice that have ethical implications for data and sample sharing in the infectious disease context are insufficiently captured in current guidance documents.

List of Tables and Figures

Tables

Table 3.1: Characteristics of interviewees 64
Table 3.2: Example of codes, categories and themes 70

Figures

Figure 2.1: Systematic scoping review flow diagram……20
Figure 4.1: Diagram of data and sample types from interview analysis.....80
Figure 5.1: Diagram of organisation types involved in data and sample sharing from the interview analysis ……114

1. Introduction

1.1. Infectious diseases as a global threat

Infectious diseases pose an important threat to global health. In January 2020, the World Health Organization (WHO) released a list of "urgent, global health challenges" (1). Stopping infectious diseases, and preparing for epidemics, are listed as two of the thirteen key challenges (1).

The WHO defines infectious diseases as those "caused by pathogenic microorganisms, such as bacteria, viruses, parasites or fungi" (2). These diseases may be passed between individuals of the same species, or between species. The spread may be directly from one individual to another, or indirectly, for example through food, water or a vector such as a mosquito. Some diseases may be endemic, and have a predictable level of circulation within a population. Others may become part of outbreaks, with a level of circulation beyond what is expected. Low- and middle-income countries (LMIC) have a higher burden of infectious diseases relative to high-income countries (HIC) (3, 4).

Infectious diseases are intertwined with human history, with some historians claiming that they have been a significant part of shaping human history (5, 6). Infectious diseases continue to pose a global threat, with particular concern around newly emerging infections, alongside known infectious diseases with epidemic potential. Endemic infectious diseases such as malaria also continue to cause significant death and disability in particular locations.

In the past 20 years, a number of infectious diseases outbreaks have captured international attention, and required international collaboration. Davies et al. in their study of international norms around infectious diseases consider the Severe Acute Respiratory Syndrome outbreak of 2003 to be an important milestone (7 p.2). They discuss how it brought to the fore the global interconnectedness that characterises our current time, and the risk that an infectious disease

outbreak in one location can rapidly spread elsewhere (7 p.2). There are a number of outbreaks in the past decade that have involved the international community, including H1N1 influenza, Middle East Respiratory Syndrome, Ebola virus disease, Zika virus, and most recently Covid-19[1] (8).

One issue brought to the fore as a result of these outbreaks is the need for international data and sample sharing to improve the effectiveness of the public health response to infectious diseases (9, 10). The hope is that such sharing would lead to a better understanding of the disease, and the development of important innovations such as treatments and vaccines. Data and sample sharing, however, also present other ethical considerations. In the next section I give a brief background to data and sample sharing, and its discussion in relation to bioethics.

1.2. Data and sample sharing, and bioethics

Data and sample sharing have received increasing attention in the biosciences and healthcare. The genomics research community is considered a pathfinder with respect to data sharing and the development of biobanks. The Human Genome Project, a large international research collaboration which ran from 1990 to 2003, established pre-publication data sharing as a norm across the collaboration. The Bermuda Principles were adopted by the project to publicly share DNA sequence data (11). The establishment of this data sharing norm among the genomic community was bolstered by the Fort Lauderdale Agreement in 2003 (12). Data sharing is now widely discussed and promoted in a broad range of disciplines in the biosciences, and in healthcare (13-17). Sample sharing, in particular the development of sample repositories known as biobanks, has grown extensively (18, 19). Large sample repositories have been developed worldwide, in order to enable a range of scientific research projects.

[1] This study was developed prior to the Covid-19 pandemic.

Both data and sample sharing have been subject to substantial scrutiny in the bioethics literature, particularly in relation to genomic research. They are characterised by their potential to provide a way of working openly, fostering collaborations and enabling science (13, 15, 20, 21). Such scientific endeavours provide a better understanding of health and disease. There has also been discussion on a range of ethically relevant issues, such as models of consent that account for uncertain future uses of the data and samples, potential risks to privacy, the potential issues with commercialisation, and the appropriate models of governance for access and use (18, 22-25). Data or sample sharing involving LMICs leads to additional considerations around equity (22, 26). For example, resource discrepancies between LMICs and HICs, can lead to scientists in LMICs losing out on academic recognition as data or samples go on to be rapidly analysed in HICs.

Data and sample sharing for infectious disease related research and responses may differ in important ways. The potential for infectious diseases to spread, and in some cases to do so rapidly, unpredictably, and widely, may mean that there are particular ethical issues that need to be considered. A systematic review of the barriers to data sharing in public health highlighted the need for more in-depth research on this topic (27).

1.3. My personal motivation to pursue this thesis

I am a public health doctor by background, and questions regarding appropriate data collection, sharing and use have come up throughout my career. There was one particular experience, however, that led me to this project. I undertook a yearlong secondment to work as a health adviser for the UK Department for International Development, in 2014-15. During this time I was deployed to the Ebola Crisis Team in Sierra Leone for a 6 week rotation as a health adviser for the UK laboratories programme. While this was clearly a short amount of time in my working career, it left a strong impression and raised a lot of ethical questions. Questions came up about sample

sharing, and the appropriate use of samples that had been collected from people presenting to the Ebola Treatment Centres. There were also questions of whether and how the data that had been collected ought to be shared. Different people appeared to have different views about what the 'right thing to do' was. There was a lack of clear guidance in the international arena at the time on data and sample sharing during an infectious disease response (9, 10).

When I returned I mulled over these questions. I had previously completed a public health training placement at the Ethox Centre, and I met with Prof Mike Parker to discuss these questions. This conversation spurred an application to undertake a DPhil to complete this project.

1.4. Thesis overview

In this first chapter I have given a very brief background to data and sample sharing, and the infectious disease context. While data and sample sharing have received a great deal of bioethical attention, this has been particularly focused on genomics. Data and sample sharing in the infectious disease context has become a more prominent topic in the wake of recent outbreaks, and the ethical issues may differ.

Chapter 2 presents a systematic scoping literature review and policy review that I completed at the beginning of this research project. This review provides an overview of the research that had been completed in relation to infectious diseases, and the gaps in the research to date. The chapter outlines the methods for the review alongside the findings.

Chapter 3 provides the details of the methods employed for this study. The chapter shows why I chose to use semi-structured interviews with professionals involved in data and sample sharing as the main form of data collection. I then present the details on data collection and analysis, before turning to reflect on my own role as a researcher.

Chapters 4 to 7 present the findings from the analysis of the semi-structured interview data. The chapters are focused on particular themes. **Chapter 4** focuses on a typology of data and samples that emerged from the analysis, and the ethical implications for sharing them. **Chapter 5** examines the different organisations and professions involved in data and sample sharing, and their varied priorities. **Chapter 6** focuses on the responsibilities interviewees consider themselves to have to others when sharing data or samples, and the role of ownership and/or control in meeting these responsibilities. **Chapter 7** goes on to explore the role and limits of consent, community engagement, trust and governance in enabling ethical data and sample sharing. The relevance of the 'infectious disease context' is a recurring thread within these themes.

Finally, **Chapter 8** examines the findings from Chapters 4 to 7, considering them in relation to the wider literature. I also reflect on the implications of the findings for policy and practice, before going on to discuss the strengths and limitations of the study, and opportunities for future research.

2. Literature and policy review

2.1. Background

In Chapter 1 I noted that recent international outbreaks have led to an increased focus on data and sample sharing to improve the infectious diseases response. Data and sample sharing are increasingly promoted in the hope of enabling more research to understand the diseases, develop innovations and improve patient care.

In order to examine the literature relevant to ethical issues arising in relation to data and sample sharing in the infectious disease context, I conducted a systematic scoping review of literature, policies and guidance documents relevant to this topic. This was undertaken firstly, to understand how and why data and sample sharing takes place in the infectious disease context; secondly, to map the ethical issues discussed in the literature and guidance; and finally to learn where there are gaps in the body of research that need to be addressed. I sought to identify research gaps in order inform the development of the study. Of note, the review was completed prior to the data collection for this study, and therefore prior to the Covid-19 pandemic which is ongoing at the time of writing. The methods and findings from this review are reported in this chapter.

2.2. Literature review: Methods

The systematic scoping review was undertaken with two questions in mind:

1. What does data and sample sharing mean in the infectious disease context?
2. What ethical issues are discussed in relation to data and sample sharing?

Systematic scoping reviews have been developed as a method to identify and map literature in a structured and transparent way (28, 29). The method is considered more flexible than more traditional systematic reviews, such as those conducted by the Cochrane Collaboration. A range

of literature types can be included, such published and unpublished studies, and conceptual literature (28-30). This is in contrast to systematic reviews, which are usually restricted to including studies with a specific methodological approach in order to answer a more focused question. The method allows iterative development of the protocol and refinement of the inclusion criteria as the review progresses. As a part of the scoping review, I also completed a review of international policy, guidance, regulatory and legal documents relating to infectious diseases data and sample sharing in the international arena. These document types will be referred to simply as 'policy' documents for the remainder of the chapter.

2.2.1. Search methods: Peer-reviewed literature

A wide range of article types were considered for inclusion from peer-reviewed journals, including empirical studies, online news articles, opinion pieces, features, reports of practice, reports of guidance, and theoretical articles. There were no restrictions based on publication date or language. Articles relating to any setting were considered for inclusion. Some articles were identified ad hoc (through newsletters, journal subscriptions or correspondence with others) and were considered for inclusion.

The literature regarding sharing public health data and biological samples in relation to infectious disease prevention, control and research was included, irrespective of whether it was part of a declared outbreak. Other disasters or emergencies may also have public health consequences, such as floods, extreme weather and earthquakes. Articles relating to data and sample sharing following such events were initially included. However, on further assessment I considered the literature about non-infectious disease emergencies to be too distinct to be included in the review. The inclusion criteria were refined accordingly to focus specifically on infectious diseases.

The review explored issues relating to data and samples collected both through research and 'routine' public health and health care activity, as infectious diseases data and sample sharing occurs in both contexts. Of note, this review did not include activities such as notification of an infectious

disease by health professionals, although it is a form of data sharing, as this review is focused on sharing datasets and samples between groups or to a repository.

The following databases were searched in October 2015:

- Medline (Ovid)
- Embase (Ovid)
- Global Health (Ovid)
- Global Health Library – Regional indexes
- WHOLIS (Virtual Health Library)

The associated keywords, MESH terms and the Medline search strategy are outlined in Appendix 1 and 2 respectively, and were developed with specialist librarian support from the Bodleian Health Care Libraries. Epidemic-prone diseases, as outlined by the World Health Organization (WHO), were included as exposure keywords (31). Global Health Library and WHOLIS were searched using free text, using the keywords shown in Appendix 1.

2.2.2. Search methods: Policy documents

Websites for organisations, purposively identified based on their potential role in infectious disease responses, were searched for relevant policy documents, guidance, legislation and regulation including: United Nations organisations such as the WHO and regional offices, the United Nations Children's Fund (UNICEF), the Office for the Coordination of Humanitarian Affairs (OCHA); the Council for International Organizations of Medical Sciences (CIOMS); the United States Centers for Disease Control and Prevention (CDC); the Department for International Development (DFID); the United States Agency for International Development (USAID); The World Bank; The European Union; European Commission's Humanitarian Aid and Civil Protection department (ECHO); the African Union; the African Development Bank (AfDB); World Medical Association (WMA); The Wellcome Trust; The UK Medical Research Council.

Médecins Sans Frontières (MSF), International Federation of Red Cross and Red Crescent Societies (IFRC), CARE International, and Oxfam websites were searched due to their size and historic roles as humanitarian health actors. Additional relevant documents were reviewed if they were referred to in the literature or suggested through informal conversations.

2.2.3. Data Collection and Analysis

I identified peer-reviewed articles for full text review, using Covidence[2] software. This software enables management of titles and abstracts, selecting to include or exclude, and is used by the Cochrane Collaboration for systematic reviews. My supervisory team assessed a sample of 20% of titles and abstracts from the peer-reviewed literature search between them, using the outlined broad inclusion criteria, for quality assurance purposes. This follows the methodology of the recent systematic scoping review by Bull et al., with this method used in order to maintain consistency of selection of literature for inclusion (28). The full-text of articles that met the broad inclusion criteria were assessed for relevance to the review question and aim. Reference lists of relevant reviews and articles matching the inclusion criteria were reviewed for further relevant literature. Papers that were not in English were either reviewed to decide on inclusion by myself (Spanish) or through colleagues providing translations, who spoke the language in question (German and Portuguese), or using Google Translate for documents (Russian). The full-text was not available for three of the articles that had been initially included on the basis of their title and abstract, following a request to the Bodleian Libraries. All papers for which the full-text was reviewed, that met the inclusion criteria, were in the English language. A flow diagram is presented in Figure 2.1.

[2] https://www.covidence.org/ (last accessed 1 October 2020)

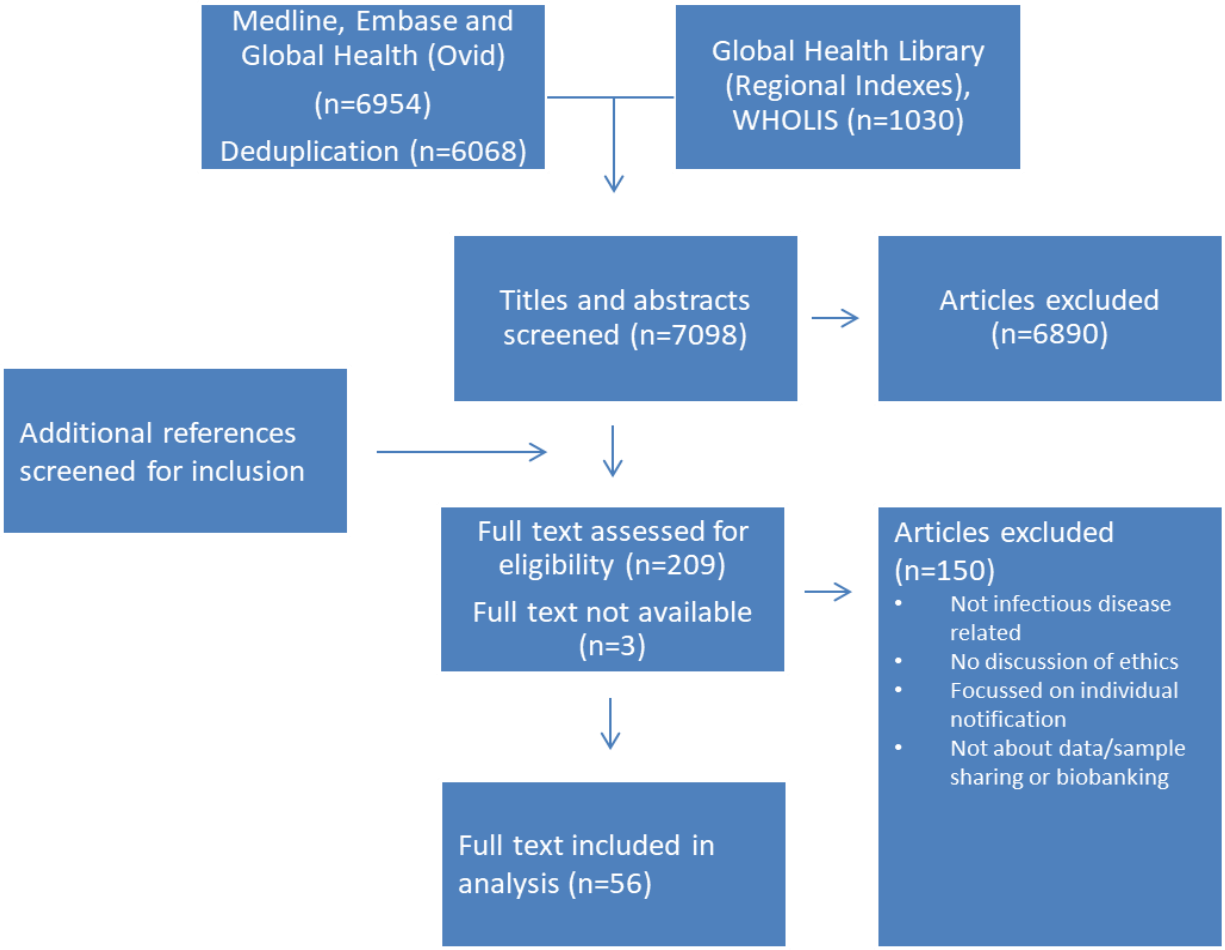


Figure 2.1: Systematic scoping review flow diagram

Of the 56 articles included after the completion of this process, 4 are empirical research papers (32-35) and 2 are workshop reports (36, 37). Two of the empirical papers are specifically related to low- and middle-income country (LMIC) contexts (32, 33). The included empirical papers are summarised here:

- Zhang et al. "Attitudes towards transfers of human tissue samples across borders: An international survey of researchers and policy makers in five countries" focused on professional stakeholders (34). They undertook a survey in China, Egypt, India, Japan, and South Korea which was quantitatively analysed.
- Patterson et al. "Gay and Bisexual Men's Perceptions of the Donation and Use of Human Biological Samples for Research: A Qualitative Study" used semi-structured telephone

interviews to examine views in relation to HIV research, and included discussion on sample sharing (35). It was based in Scotland.

- van Schalkwyk et al. "It's for a good cause, isn't it?" explored views of South African TB research participants on sample storage and re-use" used semi-structured interviews (32).
- Wendler et al. "Research on Stored Biological Samples: The Views of Ugandans" used a quantitatively analysed survey with parents of children enrolled in malaria trials to understand attitudes towards research on stored samples, and sample sharing (33).

Four of the included papers refer specifically to the Spanish HIV biobanks (38-41) and three to the King's College London Infectious Diseases Biobank (42-44). A number of the papers discuss particular events such as influenza, Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), Ebola virus and Zika virus. One paper is from Foreign Affairs (45), which is not a peer-reviewed journal. However, given its clear relevance to the topic it was included in this review. Other literature, such as books and organisational reports are discussed where relevant to identified themes.

I undertook a thematic analysis of the included literature, coding sections related to the two questions set out above. Data were managed using *NVivo 11*.

Policy related documents were mapped and summarised to examine underlying norms and principles related to data and sample sharing, with a particular focus on discussion related to infectious diseases.

2.2.4. Limitations of the review methods

It is likely that some papers and policies relevant to the topic area were not captured in this review given the breadth of disciplines relevant to the area, some of which may not be fully captured in the databases. For example, there may be additional relevant literature from databases relating to the law, to defence or international relations. In addition, for the peer-reviewed literature, not all

titles and abstracts were double-screened, but 20% of them were, for consistency, as outlined above.

2.3. Scoping review: Findings

The findings of the scoping review are presented in this section (2.3). These findings are based on themes that emerged through the analysis of the literature and review of the policy documents. The first part of this section focuses on understanding data and sample sharing in the infectious disease context. This includes an outline of how sharing takes place, and between whom (section 2.3.1), followed by a discussion of reasons that are given to share data and samples, and concerns raised (sections 2.3.2 and 2.3.3 respectively). I then outline the ethical principles, and ethical considerations that emerged within this literature (section 2.3.5). I then end this section with a more in-depth review of the key policy documents, to better understand stated norms and requirements.

2.3.1. What does sharing data and samples mean in the infectious disease context?

2.3.1.1. Who is sharing

In the literature, there were a range of stakeholders mentioned in relation to data and sample sharing in the infectious disease context. These include:

- Participants, patients
- Health care institutions
- Research institutions
- Public health institutions
- Government agencies including Ministries of Health
- Commercial organisations

The word institution or organisation captures a range of people working in various roles. These different stakeholders will have different remits, motivations and structures. In the literature there was some limited discussion on sharing data for non-health purposes, such as military, criminal justice and welfare purposes among others (46-48). However, most of the discussion focused on sharing for health-related purposes. In addition, there was some discussion of the sharing of non-health data for health purposes, for example, sharing of mobile phone data or social media data for emergency response and research purposes.

Sharing may take place at the local, national, regional or international levels (33, 34, 36, 48-51).

2.3.1.2. What is shared?

In the literature, a large range of types of data and samples were mentioned (9, 33, 34, 36, 37, 44, 49, 50, 52-67). These include, clinical data, surveillance data, pathogen genomic data, trial data, qualitative data, quantitative data, raw or analysed data. Some papers focused on sharing research data, whilst others focused on sharing of routinely collected data, some of which may then subsequently be repurposed for research. Similarly, the term 'biological samples' was used in a way that incorporated a range of sample types such as saliva, urine, blood, and foetal tissue, in addition to veterinary or microorganism samples.

2.3.1.3. How does sharing take place?

The publications reviewed described several ways in which data and samples are shared (9, 10, 32, 34-39, 41-44, 48-59, 61-78). They may be shared within defined collaborations, through informal networks, to repositories which may be open or managed access, or in the case of samples creating a biobank which may involve transfer of samples elsewhere. There may also be routine data sharing through surveillance networks of varying geographical scales.

There are numerous technological considerations related to data and sample sharing identified in the literature. One issue was the interoperability of different databases that can affect how data sharing occurs (79). In addition, some data sharing may occur in a way that does not allow best

use of the data (79-81). Merson et al. used the term "data dumpsters" to refer to "repositories of data without the metadata, data dictionaries, or documentation needed for meaningful or correct reanalysis" (80).

Commentators noted that the physical nature of samples places limits on how sharing can occur. This means that either the samples need to be physically moved to the researchers who wish to use them and stored out of the place of origin, or those looking to use them for research need to move to the samples (34, 52, 55, 78).

2.3.2. Reasons given to share data and samples

The most commonly cited reason for sharing data and samples was around maximising outputs or outcomes (9, 10, 35-37, 41, 46-50, 52, 54, 56-59, 61-70, 74-76, 80, 82-86). This may be through improving health care and public health – through planning, response and evaluation; through improving or increasing research and reducing duplication of research, with more efficient use of funds, predominantly discussed in the scientific research domain; and, relatedly, in order to achieve global health security. Some papers went on to describe sharing as an act of solidarity, suggesting that it is a recognition of our common humanity and common vulnerabilities (50, 51, 63, 72, 84). With respect to international relations, sharing can have a diplomatic impact. During the SARS outbreak of 2003, for example, China was criticised by the international community for delayed information sharing (7 p.44-72).

Other reasons were also given for sharing data and samples. One was the potential of sharing to reduce the need for additional data and sample collection and the associated harms to future participants, patients and the public (35, 37, 62, 64, 65). Pooling data can also increase the sample size for diseases or populations that have more limited research related to them (80). Lastly, some papers discussed how sharing provides the opportunity for scrutiny of results that are published and an opportunity to check analyses, thereby serving a quality assurance function (50, 53, 65, 87). A WHO training manual on ethics in public health emergencies used the case of the antiviral

oseltamivir (trade name Tamiflu) to discuss data sharing for scrutinising results (88 p.151). Governments were encouraged to stockpile and use oseltamivir for influenza on the basis of manufacturer sponsored trails, most of which were not published. A subsequent review by Cochrane of the effectiveness of oseltamivir for influenza treatment highlighted the unavailability of the data for a number of the trails, which cast doubt on the effectiveness of the drug.

2.3.3. Concerns raised with sharing data and samples

Although it is clear from the above that several reasons were discussed for sharing data, a number of concerns were also raised in the literature around sharing data and were given as reasons not to share. These concerns are at the individual patient or research participant level, institutional and global levels. Each of these is described in more detail below.

2.3.3.1. The individual patient or research participant level

At the individual level concerns raised in the literature include the risk to privacy and confidentiality which was often discussed together with anonymisation (9, 33, 35-37, 48-50, 52, 55, 57-59, 62, 64, 65, 72, 82, 85). Fairchild et al. noted that even with so-called perfect anonymisation at the individual level there remains the risk of stigmatisation of particular groups (48). Some papers suggested that people who are cases in an outbreak are vulnerable, and so these data or samples are particularly sensitive (9, 59, 72). An additional theme that emerged in the analysis of the literature was around benefit sharing with the community from which the data or samples originate, and the concerns about exploitation of the people from whom data or samples were collected (32, 34, 36, 37, 45, 46, 48, 50-53, 55, 56, 59, 61, 63, 69, 72, 74, 75, 78, 83, 84, 86).

2.3.3.2. The professional level

At the professional level (the groups amongst whom sharing is taking place), some commentators raised concerns about reciprocity in data sharing and the risk free riders using the data without sharing in turn (50, 68, 73). A related concern was the recognition of those who share, given the intensive efforts put in to collecting data or samples (9, 34, 37, 49, 50, 52, 61-65, 68, 73, 76, 80, 82,

85). Some papers raised this concern particularly in relation data or samples that have been collected in LMICs, where those in high-income countries (HICs) have more resources to rapidly analyse the shared data or samples (34, 37, 62, 82). Calain et al. specifically in reference to samples for filovirus (Ebola and Marburg) haemorrhagic fevers (FHF), noted that the safety requirements can lead to effective monopolies on where samples can be stored, with implications for who is able to decide on sample sharing (72).

In the emergency setting some authors suggested there is a tension between the speed of sharing, and quality assurance of shared data (64, 65). An example of this is provided in a Chatham House report on overcoming barriers to data sharing, which described an Escherichia coli O157 outbreak in Germany that occurred in 2011, where rapid data release from an environmental investigation was undertaken that suggested a source of the outbreak in Spain (79). This was subsequently found to be incorrect following further laboratory and epidemiological studies. By this point, however, there had been negative ramifications for Spanish agricultural exports.

2.3.3.3. The global level

In this literature concerns discussed at the global level include the challenges of reconciling the divergent interests of different groups or states. At times this was presented as a divergence between HICs and LMICs, particularly around the idea of exploitation and biopiracy[3], and the inequitable distribution of burdens and benefits with respect to sharing data and samples (36, 51, 52, 56, 63, 78, 82, 84) (see also Box 2.1). Some papers raised concerns about the potential risks of bioterrorism related to the use of shared data and samples (47, 60, 72, 89). Aarestrup and Koopmans, raised a further concern about public panic in the face of outbreaks. They suggested

[3] According to the Action Group on Erosion, Technology and Concentration (ETC Group) 'biopiracy' "…refers to the appropriation of the knowledge and genetic resources of farming and indigenous communities by individuals or institutions that seek exclusive monopoly control (patents or intellectual property) over these resources and knowledge" (for further information see: https://www.etcgroup.org/issues/patents-biopiracy - last accessed 1 October 2020).

potentially using closed groups for sharing "for public health to have time to evaluate potential undesired effects such as public panic or misuse of data" (49).

2.3.4. Ethical principles and considerations for practice

The discussion above shows that in this literature a number of reasons are given in favour of data and sample sharing in the infectious disease context. However, there are concerns too, which also have an ethical dimension. This sub-section focuses on the ethical principles and considerations for data and sample sharing practice that were discussed in the literature.

2.3.4.1. Ethical principles

A number of underlying principles for data and sample sharing practice were discussed in the literature. These include: respect for persons, patients and participants (48); fairness, equity and justice (48, 59, 61, 63); beneficence (48, 61); non-maleficence (59); dignity (9, 52, 59, 63); solidarity (50, 63); and reciprocity (50, 73).

Fairchild et al., in their paper on the use of identifiable information in the context of HIV in the United States, discussed respect for persons as a principle that underpins the need to balance privacy considerations and "use the least intrusive measures to accomplish the goal of protecting the public" (48). They also state that "just as beneficence and justice may require the use of data to limit threats to the public's health, they also demand the minimization and fair distribution of the burdens that may emerge when vulnerable populations are identified as being at increased risk" (48). This suggests that justice and beneficence are principles considered to both promote and constrain data sharing in public health.

Langat et al. discussed solidarity and reciprocity as the principles for data sharing in public health emergencies (50). With respect to solidarity they stated that:

> *[a] deadly global pandemic, or comparable public health emergency, leaves almost no lives untouched. Such dramatic events bring to the surface the deep interconnectedness and interdependence that we share as human beings. It is in such times that solidarity, that bond that integrates individuals into a people, a society and*

> *a community as a greater entity than a mere collection of persons, is put to the test. The magnitude of the threat that something like [a] global pandemic represents demands a common and unified response* (50).

This emphasises the shared connection between people globally, and its relevance as a justification for data sharing in emergencies. Reciprocity was discussed in the academic sphere, with respect to appropriate recognition of data sharing (50).

Considerations for best practice were highlighted across the data and sample sharing process, from collection, to the act of sharing, to secondary use. The most commonly discussed considerations in this literature were consent, privacy, ownership, benefit sharing, trust and appropriate governance, which I present below.

2.3.4.2. Consent

Consent was discussed in some papers in relation to respect for autonomy and the right to self-determination, and as a way of respecting bodily integrity and dignity (52, 55). Consent was also linked to maintaining trust (36). Some challenges with obtaining consent, and different consent models were discussed.

An identified challenge for informed consent was obtaining consent for future uses of shared data or samples that were not anticipated during collection, and some papers stated that it may be impractical or undesirable to re-contact participants for each use of data or samples (36, 52, 58, 82). A workshop report by MacQueen and Alleman on collection, storage, and testing of human biospecimens in HIV research suggested the need to separate consent for sharing and future uses of data and samples from consent for the study that is currently taking place, so that those who are not happy to share can still participate in the study in question (36). The report also suggested that a lengthy consent process for sharing could raise suspicions among participants that it is something concerning.

Calain et al. raised a further consideration in their paper on FHF research and response (72). They stated that "[t]he notion of informed consent…covers a range of complexities, which are usually

beyond the immediate understanding of foreign researchers working in emergencies" (72). This relates to the idea that there differing social norms with respect to acceptable data and sample sharing and use (9, 33, 35, 36, 50, 52, 72), and potential sensitivity of some samples types (33, 35, 36, 49, 57, 59) which those obtaining consent may not be fully aware of.

Various models of consent were discussed in the literature. Some commentators highlighted broad consent models as way of dealing with uncertain future uses, particularly for low risk research (32, 33, 52, 59, 78, 82). Where broad consent is used, papers discussed the role of research ethics committees (RECs) or Institutional Review Boards (IRBs) to review the proposed research on the shared data or samples. In an empirical study conducted in Uganda by Wendler et al. the acceptability of broad consent was high amongst participants, with 95% agreeing to a process of IRB approval for future uses of samples collected during a paediatric malaria trial (33). Williams et al. discussed the use of "open consent", with the option to withdraw consent in the future (44). However, two empirical studies by Patterson et al. and van Schalkywk et al. stated there was not consensus among their participants on consent models, with some stating a preference for explicit consent for each use (32, 35). Another model of consent discussed was tiered consent where participants are provided with a range of options – broad consent, consent by RECs, opt to be re-contacted prior to each use, or consent to certain diseases specific research (82). Conditional consent was also discussed briefly – a model where certain types of future uses are prospectively consented to, or not (52). Community consent or permission was mentioned only briefly (58, 78).

2.3.4.3. Privacy

Maintenance of privacy was stated to be a duty (55), linked to respect for persons (48), and needed to prevent stigma and discrimination (35, 55, 82). It was identified as an important part of the risk assessment for data and sample use (48, 52).

The paper by Fairchild et al. on HIV and identifiable public health information in the US context discussed tensions between public health and personal privacy, stating that "[t]he mission of public

health to safeguard the common good needs to be recognized but subjected to some limits by the right of individuals to a sphere of privacy" (48). Modjarrad et al. suggested that while privacy concerns are valid, in an emergency there are much greater risks to individuals and populations from not sharing (37).

Maintaining confidentiality was often discussed as one way of protecting the privacy of the people from whom data or samples were collected, either through full anonymisation, or through developing codes that are stored separately, that could allow re-identification. Some papers linked anonymisation of data and samples to making them permissible to share, where consent had not been sought (55, 58, 59, 72, 82). The possibility of retaining methods of re-identification of participants was discussed as useful if a participant withdraws to allow their data to be removed (82), or if there is a need to communicate to them at some stage (33), for example, regarding future uses. The need for an agreed upon method of anonymisation was stated by Vallance et al. (65).

Some challenges were noted, which could be categorised as technical, cultural and ethical. Technical issues related to the risks of re-identification due to limitations of anonymisation methods, although Gasner et al. suggested that the technical capacity for sharing patient data while protecting privacy was better than ever (57). A cultural challenge highlighted in the workshop report by MacQueen and Alleman was that in some contexts confidentiality may be considered a family issue rather than an issue for the individual alone (36). An ethical challenge was the risk of group level stigma and discrimination, even where data are anonymised at the individual level, or presented in aggregate form (48), and as a result such a risk remains even where individual privacy concerns are respected through confidentiality.

2.3.4.4. Ownership

Ownership of data and samples was raised in a number of papers. Ownership was generally recognised as a difficult issue, and there was a lack of consensus over who owns the data or samples (9, 10, 35). Empirical studies by Patterson et al. and van Schalkywk et al. suggested that ownership

was an area of concern among those donating their samples, with some participants continuing to identify the sample as their own or even a part of themselves, while others suggested that they no longer have ownership, or that it is a gift (32, 35). A participant in the research by Patterson et al. identified DNA as a marker of ownership (35).

Langat et al. noted that, while data collection involves time and effort on the part of the researchers undertaking it, and as a result they may feel ownership over the data, the financial investment is often made by the public, so it could be considered "common property shared among equal partners" rather than researchers being "exclusive possessors" (50). Some papers discussed the use of custodianship or stewardship models rather than ownership (37, 41, 54, 62). In these papers the suggestion was that the data or samples belongs to either the person that the data or samples were collected from (37) or the country in which the data/samples were collected (54) and that those managing the repository had a caretaker type role (41).

The concept of ownership was also discussed in relation to intellectual property rights, that is, property rights over the fruits of research on shared samples. Andanda, when discussing sample sharing in the context of a collaboration for HIV vaccine development, noted the dispersed ownership of different aspects of the process (52). The paper stated that ownership in relation to innovations developed through sample sharing can be considered distinct from the ownership of the samples themselves (52). Sharing of the fruits of innovation then becomes an issue of distribution of benefits, which is discussed in the next sub-section (2.3.4.5).

Ownership of the pathogen samples was also raised as an issue. Some publications discussed intellectual property rights on the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) pathogen gene sequence (9, 45, 83). In this case, a sample of a then unknown virus was sent from a researcher in Saudi Arabia to the Erasmus Medical Centre in the Netherlands. Following isolation of the MERS-CoV, Erasmus scientists claimed intellectual property rights on this pathogen gene sequence. This raises the question of the use of intellectual property in such scenarios. Yozwiak et

al. commented that this was a hindrance to subsequent access to the samples whilst Fidler claimed that it was in keeping with norms, to claim property rights as a result of the work taken to isolate the sequence (9, 45).

2.3.4.5. Benefit sharing

Benefit sharing was a commonly discussed point, in two main ways: the distribution of benefits with the community from which data and samples originate; and the distribution of benefits through recognition and rewards among academics or others undertaking the sharing.

With respect to distribution of benefits with the community, the most widely discussed case in the literature was that of Indonesia withholding H5N1 Influenza samples, summarised in Box 2.1 below.

In 2006/7 Indonesia withheld H5N1 Influenza samples from the (then named) WHO Global Influenza Surveillance Network (GISN) in 2006/7 (34, 46, 51, 56, 61, 63, 72, 74, 75, 83, 84, 86). This was justified by them in relation to the issue of distribution of benefits at the global level (56, 63, 75, 84). More specifically, Indonesia stated that samples that they had provided to the WHO had been transferred to an Australian pharmaceutical company without their permission (56). Indonesia's concern was that countries provided samples to the system, and pharmaceutical companies, generally in HICs developed patented products such as vaccines using these samples, which LMICs were then not able to afford, resulting in the accusation of exploitation (56, 63, 75, 84). Provisions in the Convention on Biological Diversity (CBD) were used to state their case (56, 63, 75, 84). Negotiations on the distribution of benefits following sharing of influenza samples resulted in the development of the Pandemic Influenza Preparedness (PIP) Framework (75). Indonesia was accused of compromising global health security through withholding the samples, whilst others stated that Indonesia and other LMICs were at risk of exploitation and biopiracy and sample sharing was a lever to negotiate on this issue (63, 74).

Box 2.1: Indonesia H5N1 influenza samples

The questions of distribution of benefits tended to be linked to discussions around patenting and commercialisation of innovations from shared samples. The MSF data sharing and research ethics review policies, discussed by Karunakara and Schopper et al. respectively, explicitly mention the need to avoid actions that render products unaffordable and the need for benefit sharing plans (59, 78).

The other sense in which distribution of benefits was discussed was in relation to academic recognition and rewards among those undertaking the sharing of data and samples in the context of the competitive nature of research careers (37, 49, 50, 61, 68). Data and samples require effort to collect (50) and there was concern where data or samples are shared from LMICs to HICs, that those in the HICs would be able to analyse, use and publish findings more quickly, and thus gain the recognition and rewards (80). This concern was reported to act as a potential barrier to data and sample sharing (37, 61, 64, 65, 67, 85). There were reports that some scientists working on Zika virus data did not acknowledge those who had openly shared the data they had used (73, 76). A survey study by Zhang et al. found that scientists generally agreed that commercial royalties that result from the development of innovations should be shared with the sample contributors (34).

Suggestions for practice include: attribution and acknowledgement of those who have shared their data/samples and collaborating with them where possible (37, 49, 50, 62, 68, 73, 76, 80); memoranda of understanding or data/material transfer agreements that outline the nature of the credit to be given (37, 52); posting of analysis plans by those sharing, so that others select a different analysis (9); and embargoes on data or sample sharing where it is not urgent to share (49). Where sharing is taking place between HICs and LMICs, some papers discussed the opportunity for capacity building (36, 37, 49, 51, 53, 54, 56, 61-63, 66, 70, 75, 76, 84, 86). This may be directly in relation to data or sample analytic methods, or another form of capacity building, for example, strengthening manufacturing capacity for health technologies such as vaccines.

2.3.4.6. Trust

Trust and trustworthiness emerged as a theme in this literature. This included trust of the population from which data and samples originate, trust between the groups undertaking the sharing, and trust at the global level between countries (10, 32, 35, 36, 52, 57, 61, 63, 67, 69, 71, 75, 77, 82, 87, 90, 91). It was discussed in two different senses. One was trust as a facilitator or something that is needed to allow data or sample sharing (10, 61, 63, 71, 77, 90, 91). In contrast, Brewer et al. stated that "…public trust in and support of science depend crucially on researchers sharing data, not merely advocating such sharing" suggesting sharing is a means of generating or maintaining trust (87). Some papers linked trust to transparency and openness (10, 32, 36, 61, 63), to benefit sharing (52, 61, 63, 75), and to consent processes (36). However, there was little exploration of what is meant by trust or trustworthiness.

Qualitative research by Patterson et al. on gay and bisexual men's perceptions of the donation and use of human biological samples for research found that participants generally viewed commercial organisations as less trustworthy (35). This echoes findings in qualitative work by van Schalkwyk et al. among TB research participants in South Africa that research by commercial organisations was not "good cause research" (32). This paper also found that participants trusted individual researchers to make decisions about future uses, and they were generally not familiar with RECs, although this was not unanimous.

2.3.4.7. Governance of data and sample sharing

Governance mechanisms, including policy, regulatory and legal instruments, were discussed at the individual, the institutional, the state and the global levels, and were related to the issues and concerns identified above. There are a number of such documents and instruments that relate to infectious diseases data and sample sharing (see Appendix 3).

An overarching consideration for governance processes was identified as balancing the promotion of science and public health through sharing with the interests of the various stakeholders involved

in sharing (36, 49, 52, 59, 62, 72, 78). Transparency was frequently discussed in relation to all levels of the governance processes (10, 36, 48, 56, 59, 63, 67, 69, 75).

Various governance mechanisms were discussed in the literature at the professional level where sharing is taking place. These mechanisms include:

- Review of proposals for use of data or samples by a REC or IRB, including where routinely collected data or samples are repurposed for research (32, 33, 38-41, 51, 52, 64, 82, 92). It was noted that RECs/IRBs are under-resourced and lack training in certain countries (51, 72);
- Scientific advisory committees to govern the scientific direction and prioritisation of research (38-41, 58);
- Data or sample access committees (59);
- Community advisory boards to gain input from the community from which data or samples originate (36);
- Material transfer agreements (MTA) for transfer of samples (34, 42-44, 52, 59, 75, 83);
- A MoU between collaborators (52, 78).

There were a number of governance challenges identified in the literature for infectious diseases data and sample sharing.

The first challenge related to navigating the number and diversity of existing policies and legal frameworks. While there is not an international framework for sharing data and samples in the infectious disease context (9, 10, 85), there are a number of policy, regulatory and legal instruments that are relevant (see section 2.3.5 and Appendix 3). Sharing may span jurisdictions, and involve diverse contexts with different socio-cultural norms (36, 48, 51, 52, 57, 72).

The second challenge noted by Calain et al. and Fairchild et al. is the difficulty that sometimes arises distinguishing between research and non-research data and samples, which has implications for the modes of oversight – particularly the involvement of RECs (48, 72).

The third challenge noted was developing a shared understanding of the relationship between data and samples. This has a bearing on the applicability of governance documents that relate to data sharing, and whether or not they are considered to incorporate samples. This was an issue that surfaced with the Indonesia H5N1 sample case – with debate around the degree to which sharing of "public health information" that is mandated as a part of the International Health Regulations (IHR, 2005) includes sample sharing (56, 63). This will be discussed more in the section on the IHR below (2.3.5.1)

2.3.5. The policy context

The findings from the literature show that there are a number of ethical considerations in relation to sharing data and samples. This sub-section focuses more specifically on examining the policy context, with a more in-depth review of relevant policy documents. This is to better understand the requirements and norms promoted around data and sample sharing by different institutions relevant to the infectious disease response, and how they relate to the ethical considerations identified in the literature above.

A number of policy, legal, guidance and consultation documents were identified in my search that are relevant to infectious diseases data and sample sharing. A map of those found for an initial review can be found in Appendix 3. Some of these documents operate at the institutional or national levels, whilst others have been developed by international organisations, including the WHO, or are what I have termed 'supranational' documents developed by a group of organisations or individuals. The documents may relate to different types of data or samples – some may relate only to trial data, genetic data, human tissue, and so on. In addition, the documents may have another primary purpose, such as research ethics, that has a bearing on data or sample sharing.

Those policy documents that are most directly relevant to infectious diseases data or sample sharing, with an international scope, are discussed in more detail below.[4] Many were developed through United Nations processes, in particular the WHO. Where relevant, literature found through the scoping review, alongside wider reading of books and organisational reports, are used to better understand the policy documents.

2.3.5.1. The International Health Regulations 2005

The IHR (2005) are an internationally binding set of regulations relating to the public health response to infectious diseases that spread internationally. They were revised from the IHR (1969), which was itself a revision of the International Sanitary Regulations (1951) (7 p.3-7, 93). They entered force in June 2007. In contrast to the IHR (1969), which was restricted to cholera, plague and yellow fever, the IHR (2005) apply to a broad range of infectious and non-infectious agents, including unknown agents, that may constitute a public health emergency of international concern[5] (93, 94).

The IHR (2005) state, among other things, "obligations on States Parties to notify WHO of events that may constitute a public health emergency of international concern according to defined criteria" (94 p.1). It is this obligation that is the most relevant part of the IHR to data sharing. The purpose of information sharing as defined in the IHR is to allow notification, assessment and response to such an event. It requires capacity building by states to meet the requirements, for example through strengthening surveillance systems and the development of National and WHO IHR Focal Points (94 p.11). As noted previously, the lack of clarity of the term "public health

[4] The present tense is used for the policy review, as these documents are 'live' and continue to be applied to practice.

[5] The IHR (2005) discusses a public health emergency of international concern as "an extraordinary public health event which is determined, under specific procedures: (a) to constitute a public health risk to other States through the international spread of disease; and (b) to potentially require a coordinated international response." This is determined by the following criteria: "(a) seriousness of the public health impact of the event; (b) unusual or unexpected nature of the event; (c) potential for the event to spread internationally; and/or (d) the risk that restrictions to travel or trade may result because of the event" (see https://www.who.int/ihr/about/en/ - last accessed 1 October 2020).

information" was brought to the fore in the case of Indonesia withholding from H5N1 influenza sample sharing (Box 2.1) – as it was not clear whether virus samples fell under "public health information" (56). Article 45 discusses treatment of personal data, requiring that identifiable health information collected or received "shall be kept confidential" while also noting that "States Parties may disclose and process personal data where essential for the purposes of assessing and managing a public health risk" (94 p.30). Biological substances are discussed briefly in Article 46 in relation to facilitating their transport (94 p.31).

The IHR (2005) also discuss travel and trade restrictions in relation to infectious disease events, with a view to preventing inappropriate and disproportionate responses (94). The potential for these restrictions can be a disincentive to share information (7 p.28-29).

High levels of noncompliance with IHR requirements is a reported issue, with LMICs noted to have issues with capacity (7 p.126-132, 95). In addition, the IHR only deal with "information sharing" for a specific purpose, and so do not discuss information sharing outside of the context of "unexpected or unusual public health events" that may constitute public health emergencies.

2.3.5.2. Global Health Security

Global Health Security (GHS) is a policy agenda of clear relevance of the infectious disease response. The WHO define global public health security as "…the activities required, both proactive and reactive, to minimize the danger and impact of acute public health events that endanger people's health across geographical regions and international boundaries" (96). Although encompassing a number of health threats, it is often linked to infectious diseases and bioterrorism (46, 95). HIV/AIDS, for example, was recognised as a security threat by the United Nations Security Council (Resolution 1308) in 2000 (97).

The Global Health Security Agenda (GHSA) was launched in 2014 due to concern about the threat of infectious diseases internationally and includes a number of organisations looking to make progress on IHR (2005) requirements (98). The GHSA is open to all countries and has

membership of almost 50 countries, as of December 2016 (99). The GHSA outlines a number of 'Action Packages'. One of the desired impacts of the "Linking Public Health with Law and Multisectoral Rapid Response" Action Package is: "Development and implementation of a MoU or other similar framework outlining roles, responsibilities, and best practices for sharing relevant information between and among appropriate human and animal health, law enforcement, and defense personnel and validation of the MoU through periodic exercises and simulations" (100).

The Global Health Security Initiative (GHSI) is a separate informal initiative launched in 2001 between a limited number of "like-minded" countries[6] to work on preparedness and response activities (101).

There is a wide range of discussion in relation to the relationship between security and health, and the policy discussions around GHS (7 p.17-30, 46, 95). The IHR are put forward as a tool for GHS, but also noted for their lack of provision for individual health security through access to health services and other products (95). Within the literature there was concern that the security narrative promotes unconditional sharing of surveillance data which is primarily in the interests of HICs, as LMICs may not have sufficient capabilities to respond to infectious disease threats (46, 102, 103). Heymann et al. discussed the importance of a "people-centred" approach to GHS, and the importance of access to health services and effective technologies (95).

2.3.5.3. Convention on Biological Diversity and the Nagoya Protocol

The CBD was used by Indonesia in relation to negotiations on influenza virus sample sharing, and distribution of benefits in 2006/7 (Box 2.1). CBD development was initiated in 1988, and was opened for signature in 1992, with recognition of the importance of conserving and sustaining biodiversity (104). Part of the objective is the "fair and equitable sharing of benefits arising out of

[6] The GHSI website states that "[t]he GHSI Ministerial Forum consists of Ministers/Secretaries/Commissioner of Health from Canada, the European Commission, France, Germany, Italy, Japan, Mexico, the United Kingdom and the United States. The World Health Organization is a technical advisor" (see http://ghsi.ca/ghsi-members/ - last accessed 1 October 2020)

the utilization of genetic resources" (105). Two supplementary protocols have been developed, one of which is the Nagoya Protocol which came into force in 2014, looking to address this aspect of the CBD.

The CBD, which is related to a broad range of genetic resources, asserts the "sovereign rights" of states over their natural resources (56, 74, 105). The primary purpose is about conserving biodiversity, that is, it is not primarily about infectious diseases. However, I have included it in this review as it was the main tool used by Indonesia when withholding influenza samples. There is a question posed by Fidler about the appropriateness of the use of this tool by Indonesia as, if anything, pathogens such as influenza are a potential threat to biodiversity, rather than being the sorts of biological resource we would want to conserve (56). A reference is made in the Nagoya Protocol to sample sharing related to infectious diseases which states: "Mindful of the International Health Regulations (2005) of the World Health Organization and the importance of ensuring access to human pathogens for public health preparedness and response purposes" (106 p.3). This is a more explicit recognition of the requirements of the IHR incorporating pathogen sample sharing.

2.3.5.4. The Pandemic Influenza Preparedness Framework

The PIP Framework, published by the WHO in 2011, was developed following a series of negotiations on pandemic influenza virus sharing and the distribution of benefits related to this (75, 107). It is not legally binding, and specifically relates to pandemic influenza only, not seasonal influenza or other pathogens (75, 107).

Fidler and Gostin noted that the case of Indonesia withholding H5N1 samples, followed by events in the subsequent H1N1 influenza pandemic raised concerns around the system for sharing viruses and the distribution of benefits such as vaccines (75). With respect to the H1N1 pandemic, Fidler stated that developed countries were reported to have "placed large advanced orders…and bought virtually all vaccine companies could manufacture", which then "left the developing world with

limited supplies" (84). Resultant negotiations between member states and the WHO lead to the development of the PIP Framework.

The stated objectives of the PIP Framework, to be pursued on an "equal footing" are: "(i) the sharing of H5N1 and other influenza viruses with human pandemic potential; and (ii) access to vaccines and sharing of other benefits" with a "fair, transparent, equitable, efficient, effective system" (107 p.6). The document outlines a number of principles, including "the sovereign rights of States over their biological resources", while at the same time stating the need for "collective action to mitigate public health risks" (107 p.3-5).

The section on the sharing of H5N1 and other influenza viruses with human pandemic potential discusses the use of a legally binding Standard Material Transfer Agreement that outlines a number of provisions, around capacity building, ownership, intellectual property, donation and reservation of products, among other things (107 p.29-36). Genetic sequence data, which the document states "should be shared in a rapid, timely and systematic manner" are noted to have "been considered sensitive by the country providing the virus" although further elaboration is not made on this point and the reasons behind this (107 p.13). Transparency in sharing mechanisms is taken as an important feature of the system.

The section on the benefit sharing system outlines the types of benefits from the system - such as antivirals and vaccines, alongside capacity building and technology transfer (107 p.15-22). Notably, there is a greater focus on the role and responsibilities of manufacturers to donate vaccines for a stockpile to be established by WHO, which has the aim of making available vaccines to developing, as well as developed, countries. In addition, it requires manufacturers to also pay for 50% of the running costs of the Global Influenza Surveillance and Response System (GISRS), which was formerly the GISN (107 p.21). Fidler and Gostin commented that the PIP does not address the role of developed countries to donate their purchased vaccines (75). This raises questions about who bears the costs, and who ought to bear the costs, in relation to benefit sharing.

2.3.5.5. WHO Consultation: Developing global norms for sharing data and results during public health emergencies

In September 2015, a number of international stakeholders gathered for a WHO consultation on the issue of data sharing – brought to the fore by the Ebola outbreak in West Africa in 2014-16 (108, 109). The focus was on research data, and preliminary results. In the meeting summary points sharing is put forward in moral terms: "…the fundamental moral obligation to share preliminary results once they are adequately quality controlled", with the WHO seeking a "paradigm shift" away from sharing limited by publication timelines (109). The summary sets up data and results sharing as the "default" (109). The risks of withholding data were put in moral terms.

Issues mentioned specifically around the emerging pathogen are (109):

- "The ethical requirements of informed consent" with respect to biosamples, from which data on the pathogen may be obtained;
- Accounting for the needs of the originating country: acknowledgement, inclusion in decision-making, and capacity development where needed;
- That "[p]atents on natural genome sequences could be inhibitory for further research and product development."

Sharing individual participant data and aggregate results are distinguished from each other with respect to privacy concerns (109).

2.3.5.6. WHO ethics guidance

The WHO has developed some ethics guidance in relation to epidemics and public health emergencies, which are summarised below. Given the key position of the WHO in relation to infectious disease responses internationally, these are important sources of normative guidance. Taking them as a whole, all of the guidance documents put a clear emphasis on the importance of sharing, while discussing the ethical considerations identified above in the literature.

Ethics in epidemics, emergencies and disasters training manual, 2015 (88)

This training manual provides materials to facilitators on a number of ethical issues in research, surveillance and patient care – for those working in epidemics, emergencies and disasters. There are specific sections on the collection of biosamples; publishing public health surveillance or routine clinical data as scientific knowledge; and "the ethical obligations of researchers, public health practitioners and publishers regarding ownership of scientific data". The manual discusses the potential limitations of the quality of routinely collected data, and the importance of acknowledging this in any publications based on such data.

Guidance for managing ethical issues in infectious disease outbreaks (110)

This WHO guidance, published in 2016, discusses a range of areas relevant to outbreak response, including surveillance, rapid data sharing and the long-term storage of biosamples. The document reiterates the obligations of states to notify events "that may constitute a public health emergency of international concern" as stated in the IHR (2005) "regardless of any negative consequences that may be associated with notification, such as a possible reduction in trade or tourism" (110 p.14). In the same section, obligations of HICs are also stated – to provide "foreign assistance" including for strengthening primary health care systems (110 p.14).

The section on rapid data sharing discusses it as "essential" during an emergency and notes that there are many types of data that may be shared (110 p.38). There is a brief mention of the need for countries to "review their laws, policies, and practices regarding data sharing" to account for key ethical issues such as confidentiality, managing incidental findings, and ownership and control.

A separate section discusses the long-term storage of biosamples. The guidance recognises the utility of biological samples in advancing understanding on the pathogen and outbreak (110 p.39-40). It also outlines risks – mentioning the particular difficulty in protecting confidentiality when there are small numbers, also stating that "even when individual confidentiality can be adequately

protected, some individuals or communities might still be uncomfortable making their biospecimens available for future use, especially if such use is not subject to community control" (110 p.39-40). This again speaks to the balance between public health, and sensitivity to individual and community perspectives. The need to build relationships, for consultation, and for trust is noted. Other considerations discussed in this area include (110 p.39-40):

- providing information to people about long-term storage, and seeking informed consent where feasible;
- including representatives from the countries involved for international sharing of samples, also stating that "[t]he international community should make efforts to strengthen countries' capacity to maintain biospecimens within their own borders";
- and the use of an MTA.

PAHO/WHO Zika Ethics Consultation (111)

This document was developed in 2016 following a meeting bringing together people from academic and public health organisations, the WHO, and the Pan American Health Organization (PAHO) to discuss issues that had been raised during the Zika outbreak, including data and sample sharing.

In this document, sharing data is grounded in "responsibility and solidarity" (111). With respect to research, it states that broad consent should be sought for future uses of data and samples, and that samples collected for other purposes could be used for research where such broad consent had been sought, or where the population was aware that such research use could occur. The need for ethics approvals for such studies is discussed – and the REC may request consent be sought where it was not initially given, or may allow a consent "waiver". The duty to share results of research is discussed, alongside, where possible, sharing of protocols, data and samples – thereby

explicitly making sharing an ethical position. Researchers and funders are also encouraged to build capacity in LMIC settings.

2.3.5.7. Joint statement on data sharing in public health emergencies

During the Zika outbreak in 2016, a statement was released with over 50 signatories, including funders, journals, non-governmental organisations and research institutions (112). The statement specifically relates to data sharing in public health emergencies, and is stated to follow on from the WHO Consultation on data sharing in public health emergencies that took place in 2015 – putting forward an "imperative on all parties to make any information available that might have value in combatting the crisis" (112). There are two elements to this statement, one relating to journals, and the other relating to funders (112):

- Content related to the Zika virus being made free to access by journal signatories. It also states that "[a]ny data or preprint deposited for unrestricted dissemination ahead of submission of any paper will not pre-empt its publication in these journals." This relates to concerns raised that researchers are reluctant to share data as it would impact upon subsequent publication.
- "Funder signatories will require researchers undertaking work relevant to public health emergencies to set in place mechanisms to share quality-assured interim and final data as rapidly and widely as possible, including with public health and research communities and the World Health Organization."

Of note, the emphasis of the statement is on research data, rather than data collected for routine purposes.

2.3.5.8. Médecins Sans Frontières Data Sharing Policy

MSF is a medical humanitarian response organisation that has been involved in responses to outbreaks in a number of LMIC settings. They have a specific data sharing policy (92). The remit

of the policy includes both research and routine data, and human biological material (92 p.4). There is a separate MSF policy on data protection being developed at the time of this review.

The policy is based on an "ethical imperative" and considers sharing and dissemination of routine and research data "openly, transparently and in a timely manner" a "responsibility" for the "greater public health good" (92 p.4). This is put forward as a stronger imperative given MSF's work with vulnerable and marginalised groups, as a way of improving evidence for such groups.

The MSF policy is grounded in a number of principles in the following areas (59, 92): Medical confidentiality; Privacy; Sensitive Data is protected; Equitable; Promote health benefits to the greater population; Local Benefit sharing; Intellectual Property and commercial use; Collaborative partnership; Efficient. Some points of note (92 p.4-5):

- With respect to local benefit sharing it is stated that they "will prioritise data which are of benefit to local communities" while also noting that "there is recognition that sharing can be with a wider community of individuals, and will not always result in benefits to the local community."
- The principles state that implementation by MSF will be done is a way that data sharing "will not put at risk or be misused against the interests of MSF patients, research participants, MSF employees and MSF organizations for political reasons, financial gain or any other reasons." This raises the question of what "against" means, and whether this incorporates research results from shared data that paint a negative light on MSF practice, such as an issue with the way care has been delivered.
- Under "Intellectual Property and commercial use" the policy outlines issues, such as benefits to LMICs, and that "[r]ecipients shall not seek Intellectual Property rights of any kind in respect of Results generated or arising out of the use of MSF Datasets without prior written consent." This raises a question – under what circumstances would such consent be given?

The policy outlines a managed access model for data sharing, as a form of "open data" (59, 92). "Data on sicknesses or epidemic outbreaks" are specifically mentioned as potentially sensitive data (92 p.8).

2.3.5.9. The International Severe Acute Respiratory and Emerging Infection Consortium Sample and Data Sharing Policy

The International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) is a network of researchers, hosted by the University of Oxford, established to provide a forum for open-access protocols and data sharing for emerging infectious diseases (113). It is also, in combination with the French National Alliance for Life Sciences and Health (AVIESAN) network, the secretariat for Global Research Collaboration for Infectious Disease Preparedness (GLOPID-R) (114). GLOPID-R is a network of research funding organisations looking to coordinate rapid mobilisation of funds for infectious disease emergencies (114). GLOPID-R has a data sharing working group (114).

The ISARIC policy (from 2014) has a focus on research data and samples (115). The consortium data and sample sharing policy, which is noted to be subordinate to other institutional policies, has an objective to "…maximise the access and sharing of samples and data among members of the ISARIC research community and external collaborators working in severe emerging infections" (115). It excludes data and samples generated through research fully funded by the private sector (115).

A few points of note from the policy (115):

- A number of principles are mentioned: respect for participants/patients through confidentiality; maximising benefits and minimising harms; increasing public respect for clinical research; and "[e]nsuring fairness for individuals and society".

- Data ownership remains with the investigator, who would be the data contributor. "Principles of Accountability" are outlined which include mention of intellectual property rights, professionalism and legal conformity.
- A number of policy recommendations are made. One states that "mechanisms should be sought to ensure that the flow of royalties is balanced including protecting the appropriate rights of participating communities." There is not further discussion on the implications of this for practice. Enhancing technology transfer to developing countries is also mentioned.

2.3.5.10. Key points from the policy review

The policy review has shown that there are a number of documents relevant to data and sample sharing, and there will be others, not reviewed here, that apply to individual countries and organisations. It is notable that policies and guidance often discuss the need for data and sample sharing explicitly in terms of ethics. Outbreaks also appear to be a moment where renewed attention is given to the need for, and potential issues with data and sample sharing. The WHO Consultation summary (section 2.3.5.5) and the joint statement (section 2.3.5.7) were developed following the outbreaks of Ebola virus in 2014-16 and Zika virus in 2016 respectively.

One further notable observation across many of these documents is that while the sharing data and samples are often discussed as essential, it is unclear how in practice this is balanced with individual, community and state interests. A number of such interests are discussed in the policy documents, and aligned with the ethical considerations noted in the literature review. This includes the need for consent, maintaining confidentiality, being sensitive to community perspectives, and considering the distribution of benefits. It is not clear how those involved in data and sample sharing navigate these issues in practice, which may pull in different directions.

Finally, it is notable that there is ambiguity in the scope of some of the policy documents - for example whether public health information includes samples, in the case of the IHR (2005). In

others, this is more explicit, such as the MSF policy, which notes samples as a type of data, and the WHO ethics guidance which, conversely, has separate sections for data and samples.

2.4. Conclusion

2.4.1. Scoping review conclusions

This scoping review shows that data and sample sharing in the infectious disease context may involve a number of different organisations, sharing different types of data and samples in different ways. There were a number of reasons given to share data and samples, and also concerns raised with sharing. These reasons and concerns have ethical dimensions. Some key ethical considerations were put forward that are often discussed in the bioethics literature in general – such as consent, privacy, ownership, benefit sharing, trust and appropriate governance.

The review suggests that the infectious disease context has some distinctive features. The first is the speed at which sharing may need to take place (noted in section 2.3.3 and also notable in the policy documents). The second is the communicable nature of the diseases, which provides a particular impetus to share. This is particularly evident in the discussion around the case of influenza virus sharing from Indonesia, and the stake the rest of the world was seen to have in those samples. There is also a sense in this literature, and some policy documents, that data or samples originating from an outbreak are more sensitive, although the rationale behind this is not clear.

Sharing from LMICs, which face a disproportionate burden of infectious diseases, also raised questions around the fair distribution of benefits. The potential global nature of the sharing across different jurisdictions and areas with different socio-cultural norms is another notable feature of the discussion in the literature.

2.4.2. Research gaps

This review shows that there had not been an in-depth study of the practical experiences and views of the ethics and governance issues encountered by those who work in roles relevant to infectious diseases data and sample sharing. As noted above, a number of ethically relevant areas were discussed in the literature. However, these were predominantly discussed in commentary pieces, theoretical articles, and policy documents, rather than empirical studies, and therefore appeared either removed from practice, or commentaries grounded in the specific experience of the author(s) recounted from their perspective. A number of questions also arose from the literature review, but were not answered by it, such as what constitutes sharing, and the relevance of what is shared – data, samples, and the sub-types - to the ethical issues arising when sharing. A light touch review of literature during the Covid-19 pandemic showed that there remains a research gap, as articles were generally commentary pieces and editorials.

Only four studies in the literature review were empirical studies (summarised in section 2.2.3). All four studies were specifically focused on samples, and three of the four studies above were focused on views and attitudes of the people from whom data and samples had been collected. Only Zhang et al. focused on the professionals undertaking sample sharing (34). This study provided important insights into researcher views on material transfer agreements, intellectual property and authorship. The study did, however, use a pre-specified set of questions, and did not provide a more in-depth exploration of the issues that study participants themselves considered important. In addition, given the focus on samples, as with all of the empirical studies in the literature review, there was no exploration of the relationship between data and sample sharing – although some policy documents discuss them as if they are interrelated.

Given there had not been any in-depth study of the practical experiences and views of the ethics and governance issues encountered by those who work in roles relevant to infectious diseases data and sample sharing, and are engaged with it, this study sought to address this research gap. By

providing a better understanding the perspectives of those in roles relevant to data and sample sharing, the study will, I hope, provide a better understanding of the ethical and governance challenges faced by people engaged with data and sample sharing in relation to infectious diseases. The methods for this study are described in the next chapter.

3. Methods

This chapter describes the methods used for this study, and the rationale behind the selected methods. The chapter starts by outlining the aims and objectives of the study, and the reasons for choosing the research methods. The remainder of the chapter is focused on describing the research methods, including data collection and analysis, before going on to reflect on ethical considerations in relation to the research, and reflections on my own role as a researcher.

3.1. Aim and objectives

Informed by the research gap discussed in Chapter 2, the overarching aim of this study is to examine perspectives of the professionals involved on the ethical issues, and the role of ethics governance, in relation to infectious disease data and sample sharing.

The objectives of the study are to examine:

a. Practical experiences and views on the benefits and challenges of data and sample sharing;
b. The concept of 'sharing', how this is understood by those working in roles related to infectious disease data and sample sharing;
c. How ethical issues in sharing are identified, characterised and resolved in practice;
d. Views about what constitutes 'good data sharing practice' in context of infectious disease responses;
e. How processes are, or could be, developed to identify and address ethical issues in these contexts.

The focus is primarily on sharing from low- and middle-income countries (LMIC) as the burden of infectious diseases in proportionally higher in these contexts, and the resource constraints are likely to present additional ethical and governance challenges.

3.2. Choice of research method

In order to meet the aims and objectives, a qualitative methodology was required, as an approach which allows the ethical issues present in the everyday experiences of data and sample sharing to be explored. In what follows, the possible research methods considered for this study and the rationale for the methods selected will be discussed.

3.2.1. Possible choices of method

Based on the identified gaps in the literature and the specific aims of this research the methodological approach taken needed to be able to elicit the perspectives of those people with everyday experiences of data sharing and sampling in the contexts of infectious diseases. To achieve this, I considered a number of different approaches and reviewed some studies in the wider field of data and sample sharing ethics and governance, to assess suitability for this study. The empirical studies identified through the literature review (Chapter 2 section 2.2.3), used either semi-structured interviews or surveys. I also reviewed a selection of other studies that I came across either during the literature search that were not included in the review (generally as they were not grounded in the infectious disease context) or through wider reading on the topic.

A series of informative studies in the Journal of Empirical Research on Human Research Ethics published in 2015 examined "stakeholder experiences of, and views about best practices in sharing individual-level data from clinical and public health research" (22). The studies also included professionals (as well as research participants) in five LMIC settings. The researchers used a mixture of semi-structured interviews and focus group discussions to collect data. The interviews were used for those who had some experience of data sharing, whilst focus groups were reserved for those with less experience and were facilitated by vignettes to aid understanding of data sharing.

These methods are often employed to understand perspectives on ethical and/or governance issues in relation to the topic in question. Several other studies exploring the views of professional

stakeholders have also employed interviews and/or focus groups. For example, Tindana et al. also used a mixture interviews, focus group discussions and observations to understand stakeholder perspectives on the collection, export, storage, and use of human biosamples at two sites in sub-Saharan Africa (26). De Vries et al. used document reviews and interviews when exploring views on addressing potential group harms from genomic data sharing (116), and Parker and Kingori used interviews to examine views of professionals in relation to good and bad research collaborations (117). These are a few of many potential examples deploying these methods, and the papers demonstrate the utility of these methods in exploring and understanding the views of those participating in the research.

Some studies explored views of professional stakeholders in data sharing and/or biobanking used surveys, such as Zhang et al. (included in the literature review) and Rathi et al. who surveyed clinical trialists about their views on sharing data (34, 118). These studies used pre-specified questions, related to areas that the research team were looking to gather views on. These surveys provided knowledge on the agreement (or not) of those participating in the study to pre-specified responses, and the proportions of people selecting different responses. These studies also have a clear role in shedding light on views and perspectives in relation to questions of ethics and governance, particularly where agreement or disagreement with pre-specified positions is being sought.

A further approach described in two of the papers included in the literature review, among others, was a stakeholder workshop to discuss the topic in question (36, 37). The approach in these papers was to convene experts in a particular area to discuss the topic in question. This provided a way of stimulating discussion, noting areas of consensus and divergence of views in relation to a topic, and allowing an exchange of ideas between workshop participants.

3.2.2. Rationale for using qualitative methods in this study

The studies outlined above employed one of two broad research methods – either quantitative surveys, or qualitative methods, predominantly semi-structured interviews and focus groups, but

also document reviews and observations. The surveys enabled the quantification of the proportions of people agreeing with particular views and statements. The qualitative studies provided insights into why study participants held particular beliefs or attitudes, how their beliefs or attitudes were formed, and an exploration of their experiences and practices. Workshops enabled discussion, and the noting of areas of agreement and disagreement, and the possibility of collectively agreed statements on a particular issue.

Given the aim and objectives of this study and following reflection and discussion with my supervisory team, I reached the decision that qualitative methods were the most appropriate method to enable understanding of experiences sharing data and samples, exploration of what stakeholders perceived to be important with respect to ethics and governance (rather than pre-specifying this), and why they hold these viewpoints.

The next question was to determine the methods of data collection. After considering the relative strengths and limitations, I decided to primarily use semi-structured interviews with professional stakeholders. As the study is focused on people with experience of sharing, or already engaged in the topic of sharing in some sense, they have familiarity with the area. Semi-structured interviews allow focus on certain topics through an interview guide, but flexibility and deviation from the stated questions to explore certain topics further (119). This seemed to be appropriate to the research objectives – as I planned to focus on some areas of particular interest with respect to infectious diseases data and sample sharing, but I also wanted to remain flexible and responsive to arising topics and emerging views that I may not have considered at the inception of the study, that are relevant to the research aim.

I decided that focus group discussions would not be appropriate for this study, as this study was not seeking to explore the interactions between those participating in the study, and given the hierarchical nature of many of the professionals involved in data and sample sharing, there was a risk that people participating in focus groups would be influenced by those considered more senior

or knowledgeable. In addition, it was judged that those participating in this study would be likely to have a professional relationship with each other, which could also constrain what is said in a group setting.

In addition to the interviews, I also chose to undertake observations of meetings related to the data sharing platforms which are discussed in more detail below. Such observations contextualised the setting in which the discussions on such platforms take place, the nature of the discussions, and the dynamic between different parties. Using more than one mode of data collection also allows a triangulation of data sources (in this case interviews and observations). Although I have not used the observation notes as explicit sources of data in the findings, these meetings provided important contextual information and influenced the development of my interview guide and areas to focus on. This is discussed more in section 3.4.3.

3.3. Research context

3.3.1. The Infectious Diseases Data Observatory – background

As noted in the Chapter 1, data and sample sharing has received increasing attention in the infectious disease context. The nature of infectious disease preparedness and response is that it transcends borders and involves people and institutions in widespread geographical locations. As a result, I did not focus on studying specific countries, but sought to identify interviewees who had relevant professional experience – specific inclusion criteria are discussed in more detail in section 3.4.2.1.

I worked with the Infectious Diseases Data Observatory (IDDO)[7], based in the Nuffield Department of Medicine at the University of Oxford to identify relevant stakeholders for the

[7] Further information on IDDO can be found at: https://www.iddo.org/ (last accessed 1 October 2020)

study. The study did not focus on IDDO as a specific case study, but links with IDDO enabled me to better understand data sharing, and to identify interviewees.

IDDO developed from the WorldWide Antimalarial Resistance Network (WWARN) collaboration. WWARN provides evidence on antimalarial medicine efficacy and resistance through conducting pooled meta-analysis on data shared from clinical trials. It was developed in 2007 through funding from the Bill and Melinda Gates Foundation, and in 2009 was formalised as a network with the secretariat hosted at the University of Oxford (120).

IDDO have subsequently been developing other data sharing platforms, including for Visceral Leishmaniasis (VL) and Ebola Virus Disease (EVD), Schistosomiasis and Chagas Disease. As a result, the platforms hosted by IDDO relate to data collected in both emergency and non-emergency settings, in different geographical locations. Members of the group are also involved in biobanking discussions related to the Ebola outbreak in West Africa in 2014/15, and others in the wider network are involved in sample sharing.

VL and EVD were two of the data sharing platforms being developed during the data collection period. I chose to interview stakeholders relevant to these platforms, and so present a brief background here.

3.3.2. Visceral Leishmaniasis data sharing platform

VL is a vector-borne disease cause by the *Leishmania* genus of protozoan parasites, and is potentially fatal if left untreated (121). The disease is endemic to East Africa, South Asia and Brazil and is listed by the World Health Organization (WHO) as an epidemic-prone disease (31). The WHO also identifies leishmaniasis as "a poverty-related disease" going on to say "[i]t affects the poorest of the poor and is associated with malnutrition, displacement, poor housing, illiteracy, gender discrimination, weakness of the immune system and lack of resources" (122).

The VL platform was developed by IDDO to pool clinical trial data related to VL treatments, to enable individual patient data meta-analysis. It is coordinated by IDDO, in conjunction with the Africa Office of the Drugs for Neglected Diseases Initiative (DND*i*)[8], which is a non-profit research and development organisation. The platform description states that "IDDO provides an opportunity for VL researchers to pool data from different VL clinical trials in order to enhance the statistical power of the available data. A large pooled dataset will be more effective in determining optimal dose ranges and will provide a deeper insight into VL characteristics" (123). The platform has a Scientific Advisory Committee (SAC) to give advice on the strategic research priorities of the platform and a Data Access Committee (DAC) who make decisions on access to the data (124).

3.3.3. Ebola Virus Disease data sharing platform

EVD is an epidemic-prone zoonotic disease with person-to-person transmission, and a case fatality rate of about 50% (125). The largest outbreak recorded so far occurred in West Africa in 2014-2016, with 28,652 cases and 11,325 deaths, mostly in Guinea, Sierra Leone and Liberia (126). At the time of writing (2019/20) there has been a subsequent outbreak in the Democratic Republic of Congo.

The EVD platform has been developed to pool clinical, epidemiological and laboratory data from outbreaks. Alongside enabling research, an objective is "to facilitate collaboration, training and capacity strengthening in Ebola-affected countries" (127). There is a steering committee, tasked with policy and strategy in relation to the platform, which includes representation from academia, public health agencies, governments and non-governmental organisations (128). Similarly to the VL platform, there is also a DAC (128).

[8] For more information about DND*i* please see: https://dndi.org/ (last accessed 1 October 2020)

3.4. Data collection

3.4.1. Data sources – an overview

The main data collected for this study were 28 semi-structured interviewees conducted between October 2016 and October 2017. My research was also informed by two other activities undertaken during the data collection phase. The first was attending and observing meetings where data sharing platforms were being discussed. The second was having regular informal conversations with people based at IDDO to better understand the platforms being developed, and to remain informed about their progress.

3.4.2. Semi-structured interviews

3.4.2.1. Sampling Frame

This study included people, over the age of 18, who had been identified as working in an area relevant to data and sample sharing in the infectious disease context. The interviewees were considered to be best placed to provide insights into the practical ethical and governance challenges associated with sharing data and samples, as they had relevant professional roles. Interviewees working in a wide range of roles were included in the sampling frame in order to develop a more holistic understanding of a variety of professional perspectives, including: data managers; laboratory managers; study coordinators; researchers; health professionals; funders; ethics committee members; in country regulators and policy makers; and international policy makers. The study did not include interviews with research participants, patients or the general public because the focus of the study, as noted in section 3.1 was specifically related to the professionals' experience of and views about sharing data and/or samples between groups or to a repository.

Interviewees were identified through two main forms of sampling. Firstly, purposive sampling was undertaken using three main sources to identify people, with a view to sampling across the professional roles noted above. Those selected for interview were:

- involved with or being engaged by IDDO with respect to data sharing;
- within my own public health networks who were involved in data or sample sharing;
- identified through reading meeting notes and newspaper or journal articles.

Purposive sampling enabled me to identify a range of prospective interviewees across the sampling frame, to ensure that a variety of different perspectives on the topic were included. From the literature review it was apparent that the sharing of infectious disease related data and samples took place between a range of professions and organisations, across and between high-income countries (HIC) and LMICs, and so I considered it important to ensure there was variation among the interviewees with respect to these parameters. It was also clear from my review of the literature that there was a lack of diversity in perspectives in previous studies of this topic, so this feature was included in the methodological approach adopted in this study.

Initial interviews were focused on VL and Ebola, as two contrasting infectious diseases. VL is endemic to certain regions, while Ebola is an emerging infection that appears in the form of outbreaks. However, during the initial interviews it became apparent that many interviewees had a breadth of experience spanning different forms of data and/or sample sharing in different contexts, for varying diseases, and these experiences were of clear relevance to this study. As a result, the study was extended to include a wider range of interviewees who worked on a range of diseases relevant to the study objectives, beyond VL and Ebola.

Snowball sampling was also used, where people who had been interviewed were invited to recommend others they thought it would be valuable and relevant to interview. This allowed me to reach people who may have diverse perspectives in the wider network who worked on relevant areas but were not necessarily known to me or attending meetings or directly interacting with IDDO, thereby increasing the reach of the sampling.

3.4.2.2. The process of recruitment

Interviewees were invited by email with the participant information sheet attached (see Appendix 4). Recruitment was primarily done through email to ensure those invited had the opportunity to read the participant information sheet in their own time prior to responding, to verify my own identity as emails were sent from my University of Oxford email address, and to reduce the social pressure to participate that may occur with an in-person invitation. All interviewees were given the opportunity to discuss the study in more detail. I had previously met some of the interviewees, and had discussions with them about the study, while others asked some questions prior to the interview formally commencing.

A small number of interviewees wanted to further discuss the expertise required to be interviewed and were concerned about not having enough background knowledge about ethics, or not being senior enough in their workplace. I was careful to explain that while they were under no pressure to participate, that they did not need any prior knowledge in ethics, or a particular level of seniority to be interviewed, and this seemed to allay this concern.

I maintained an Excel spreadsheet, kept confidentially in an encrypted document on the computer drive at the Nuffield Department of Population Health, to keep track of who had been invited, who had been interviewed, who had declined and who had not responded. Four people formally declined to be interviewed because they felt it was not their area of work (but did in some cases suggest others who they considered appropriate), twelve people did not respond to the request at all, and eight people were unable to find an appropriate time after initially accepting the request.

Informed consent sheets were completed by the interviewees, with explicit consent for audio-recording of the interviews. The consent form can be found in Appendix 5.

3.4.2.3. Conducting interviews

Interviews were conducted face-to-face where people were based in England, and otherwise were conducted through video calling where possible, or audio calling where the video calls were not

working adequately. Further details and reflections on the possible differences between face-to-face and remote interviews are presented in section 3.4.2.4 below. Two initial pilot interviews were completed with people previously known to me who met the inclusion criteria. These pilot interviews enabled me to practice my interview technique, and road test the interview guide. I sought feedback from the interviewees about my style and any suggested areas for improvement. One area I noted, and we discussed, was my tendency to summarise and reflect on what the interviewee had said, introducing the risk that I was at times putting words into their mouths. I refrained from doing this in the subsequent 26 interviews, aside from seeking clarifications. As these interviews were also an important data source providing important perspectives relevant to the research objectives, and changes resulting from the pilots were very minor, they were included in the final analysis.

The interviews lasted between 28 and 93 minutes, and were concluded when there were no additional questions on my part, after following up on interviewee responses and consulting the interview guide, and the interviewee felt that they had nothing further to add. Four interviewees gave a specific time limit to the interview as they had a number of meetings on the same day, and the interview was therefore concluded at the end of the allocated time. All interviewees agreed to be audio-recorded.

Characteristics of the 28 interviewees can be found in Table 3.1. A limited set of characteristics are presented to maintain confidentiality, and information on both stated nationality and where the interviewee was usually based (at the time of interview) have been grouped by continent for the same reason. These characteristics included here were considered, following data analysis, to be the most relevant to interpreting the interview data.

As shown in Table 3.1, interviewees from a range of professional backgrounds were invited to participate in this study, with some interviewees having two different backgrounds – such as being both a doctor, and a policy maker or researcher. Some interviewees were also employed by more

than one organisation. The majority (n=18) worked in research institutions, 5 worked for a government agency, including funding agencies, 1 additional interviewee worked for both a government funding agency and a philanthropic funder, 2 worked for non-governmental organisations and 2 worked for regulatory agencies. The majority of interviewees were based in the same continent as their stated nationality, with 13 based in Europe, 10 in Africa, 3 in North America, 1 in Asia and 1 in South America. While I have not specified this for each individual (for reasons of confidentiality), across the interviewees, people who have experience in relation to a range of infectious diseases were represented, including VL and Ebola, alongside HIV, acute respiratory infections, malaria, viral hepatitis, and trypanosomiasis. Some of these diseases, such as VL and malaria are endemic in certain parts of the world, while others, such as Ebola, and certain acute respiratory infections, present in the form of outbreaks.

The design of the interview guide was informed by the systematic scoping literature review presented in Chapter 2, and can be found in Appendix 6. Areas of discussion in previous interviews were used to inform subsequent interviews and further explore arising themes.

Table 3.1: Characteristics of interviewees

Interview	Profession	Stated nationality	Where based	Type of institution
1	Laboratory Scientist	Europe	Europe	Government agency
2	Epidemiologist, Researcher	Europe	Europe	Research institution
3	Policy Maker	Europe	Asia	Government funding agency
4	Pharmacologist Researcher	Europe	Europe	Research institution
5	Clinical Trials Manager	Africa	Africa	Research institution
6	Doctor, Research Director	Europe	Europe	Research institution
7	Clinical Trials Manager	Europe	Europe	Research institution
8	Clinical Trials Manager	Africa	Africa	Research institution
9	Pharmacologist, Principal Investigator	Africa	Africa	Research institution
10	Research Ethics Committee Member, Social Scientist	Africa	Africa	Research institution
11	Doctor, Principal Investigator	Africa	Africa	Research institution
12	Doctor, Laboratory Director	Europe	Europe	Government agency
13	Statistician	Africa	Africa	Research institution
14	Doctor, Policy Maker	Europe	Europe	Government funding agency
15	Doctor, Scientific Director	Europe	Africa	Research institution
16	Scientific Director	North America	North America	Research institution
17	Clinical Trials Researcher	North America	Europe	Research institution
18	Doctor, Policy Maker	Europe	Europe	Government funding agency
19	Data Manager	Africa	Africa	Research institution
20	Epidemiologist, Researcher	Europe	South America	Research institution
21	Public Health Doctor	Africa	Africa	Government agency
22	Doctor, Principal Investigator	Europe	Europe	Research institution
23	Doctor, Adviser	Europe	Europe	Non-governmental organisation
24	Public Health Worker	North America	North America	Regulatory agency
25	Doctor, Government Worker	Africa	North America	Regulatory agency
26	Researcher, Adviser	North America	Europe	Non-governmental organisation
27	Sociologist, Researcher	Oceania	Europe	Research institution
28	Doctor, Scientific Director	Africa	Africa	Research institution

3.4.2.4. Reflections on face-to-face and remote interviews

In this section I present some reflections on remote (video or audio) interviews, as compared to the more usual face-to-face interview. As noted in Table 3.1, interviewees were geographically dispersed. The nature of data and sample sharing is reflected in this study, in that those who are stakeholders are part of geographically dispersed networks. Using video or audio calling allowed interviews to take place which would have been very difficult to arrange face-to-face, particularly as interviewees were often travelling. The decision to do interviews outside of the UK remotely through video or audio calls was primarily based on methodological considerations, allowing a wider reach and a resultant inclusion of a broader range of interviewee perspectives, in order to better meet the research objectives.

It is important to consider the potential benefits and drawbacks associated with remote interviewing. Alongside the benefit that it enables interviews to take place with geographically dispersed people at a time of their convenience, there is the potential for remote interviews to be more convenient for research participants, as well as the researcher, to allow the interviewee to be somewhere that they are comfortable, and to allow the interviewee to terminate interviews more easily at any moment that they choose to do so (129-131). Potential drawbacks noted with remote interviews are the technical issues disrupting the interview, the challenges of reading non-verbal cues and developing a rapport between the interviewer and interviewee (129-131).

I found that it was not straightforwardly the case that conducting an interview face-to-face, which can be seen as the default mode of interviewing, was always better than using video or audio calls. I found that the rapport varied from interview to interview and it did not feel at the time, nor subsequently appear in the transcripts, to be clearly related to the mode of interview. Face-to-face interviews lasted between 55 and 93 minutes and remote interviews lasted between 28 and 83 minutes. Some interviewees that were speaking remotely were in their own homes, and in fact seemed more relaxed compared to those I was interviewing face-to-face in their workplace.

Conversely, I did find it helpful to better understand an interviewee's work context by visiting their offices for an interview, and it was usually easier to strike up a more casual conversation before initiating the formal part of the interview. I did experience some technical interruptions to the video and audio calls which at times disrupted the flow. I was also aware that the audio recording was more hidden for those being interviewed remotely, which is also discussed by Lo Iacono et al. (129), and so I always ensured that the interviewee knew when the recorder was being switched on and off.

Although the decision to interview remotely was made on methodological grounds, an additional benefit of the remote interviews was a reduction in the carbon emissions, and the financial expense associated with this study. The costs associated with face-to-face interviews at long distance in terms of carbon emissions, money and time are notably different to those associated with calls – and this has wider implications for the sustainability of research. Given the climate emergency is widely recognised as a grave threat to human health and wellbeing, I believe the greenhouse gas emissions associated with the research study are an important area to reflect on. The Tyndell Centre for Climate Change Research states that academic researchers are "high emitters" as a group, attributing this primarily to flying – including for fieldwork (132). Whilst using technology for audio and video calling is not a carbon neutral activity, it has far lower carbon emissions than those associated with travel, particularly by aeroplane. I did travel for the meetings discussed in section 3.4.3 which allowed me to meet some people face-to-face to develop an initial rapport, who I then went on to interview remotely. I also reviewed the initial video and audio interviewees to ensure there were not marked disparities in the quality of the discussions prior to continuing to conduct further interviews remotely.

3.4.2.5. Completion of interview data collection

Twenty-eight interviews were conducted in total, including the two initial pilot interviews that on review were deemed suitable for inclusion in the analysis. My supervisors and I regularly reviewed

the process of data collection and analysis together, and I ceased data collection when we reached agreement that limited additional insights were being gained. No new codes were created following on from interview 24, and a decision was made to suspend further interviews following interview 28. There are some groups that remain under-represented in relation to the sampling frame, particularly policy makers in LMICs, and WHO staff, although several attempts were made to contact individuals from these groups. As such, although theoretical saturation appears to have been reached, it is possible that new perspectives may have been gained through interviews with individuals from these groups.

3.4.3. Meeting observation

In addition to informal conversations with those based at IDDO, I attended some more formal meetings to better understand the context of data sharing. Sample sharing was not a focus of these meetings, although some people who were also involved in sample sharing were in attendance. These meetings also gave me the opportunity to identify and meet with potential interviewees.

I attended:

- The Leishmanisis East Africa Platform (LEAP) meeting in Nairobi, Kenya in October 2016
- The IDDO co-ordination meeting in Oxford, UK in December 2016
- World Leish 2017 in Toledo, Spain in May 2017

At each of the meetings I was given the opportunity to introduce myself and my research, listen to presentations and discussions around the development of the data sharing platforms, and learn about ongoing research in the area. Those attending the meetings were usually researchers, but there were sometimes others present, such as ethics committee members and data managers.

Through attending these meetings, I was able to gain insights into both support for, and concerns raised, about the development of the IDDO VL data sharing platform described briefly in section 3.3.2, and a wider sense of the dynamic among the researchers who were being engaged to share data. It was also then an opportunity to purposively invite those who had aired different views with respect to data sharing to be interviewed. I did keep notes, without including individual names, on the discussion at the meetings - being particularly attentive to areas of discussion relevant to the study objectives, and any points of apparent consensus, or conflict and disagreement. I have not included direct references to the meetings in the analysis as permission was not sought from those attending to use the meetings in this way. However, attendance has informed my knowledge of the subject area, and provided me with insights of particular areas of contention to follow up during interviews, such as the perceived role of research ethics committees in data and sample sharing, and views on the need for consent from the individuals from whom data or samples were initially collected.

3.5. Data analysis and presentation of findings

3.5.1. Approach to analysis

Audio recordings of the interview were sent for transcription by an external transcription service[9] which was subject to a non-disclosure agreement. On receipt of the written transcript, I verified it through listening to the audio recording while reading the transcript. I was able to correct errors, refresh my own memory on the tone of voice and other aspects of speech that would inform my understanding analysis of the transcript, and get a global sense of the areas discussed in the interview.

[9] The company was https://www.rev.com/ (last accessed 1 October 2020)

I then undertook a thematic analysis to identify, organise and interpret patterns that were found in the data. The 6-phase guide described in Braun and Clark's paper "Using thematic analysis in Psychology" was used as a basis for this analysis, as well as being informed by the description of thematic analysis given in Bryman's "Social Research Methods" (119, 133). I decided that thematic analysis was the most appropriate method of analysis as I was committed to a data-driven, inductive, approach to the analysis, with the findings being grounded in the perspectives of the interviewees (although I acknowledge I cannot claim to be impartial, and reflect more on my own role as a researcher in section 3.7). In order to meet the study objectives, I planned to synthesise findings from across the interviews, and as thematic analysis is also focused on patterns across the dataset rather than individual interviews, this is a well-suited analytic method to this study (133). I did also consider framework analysis as it provides a clear way of structuring the data which appealed to me, but this did not seem as well suited to the inductive nature of the analysis for this study, and I and my supervisors felt that it was better suited to studies where there was more of a deductive element to the coding process.

I started the analysis with line-by-line open coding of the interviews, developing inductive codes grounded in the interview data. The list of codes, with their description, can be found in Appendix 7. The initial codes that were developed were then grouped together into broader categories, encompassing multiple related codes – and these categories from the basis of the themes discussed in the findings (example provided in Table 3.2). The codes that were used as building blocks for the themes discussed in this thesis were those that related to the research aims and objectives, and were either frequently discussed by interviewees, or were areas of contention either for an interviewee individually or between interviewees. I used the 'one sheet of paper' technique described by Ziebland and McPherson to aid visualisation and interpretation of the breadth of data in a code (134). The technique involves mapping the content of a code alongside the interview numbers on a single sheet of paper, to get a sense of contrasting perspectives in relation to that code. Differences in perspective were more easily seen in this way, for example under the code

"data vs samples" I had mapped the content under sub-headings such as "not different", "data fixed", "samples physical", "genomic especially sensitive" and so forth, to visualise the content of the code more clearly. During the analysis I continued to refer back continuously to the original transcripts, as well as using the sheets of paper to help provide structure and look at relationships between codes, to build them into categories.

Table 3.2: Example of codes, categories and themes

Initial code	Broader category	Themes
Data vs samples Sensitive data Research vs routine Qualitative vs quantitative Confidentiality and anonymisation	Data and sample types	Data and sample typologies, and relationship to ethical sharing
Definition of sharing Sharing with a team International sharing Keeping in country Repositories vs collaboration Open data	Types of sharing	
Value of sharing Privacy Confidentiality and anonymisation Good practice Bad practice	Sharing different data and sample types	
Infectious disease context Different infectious diseases Stigma	Sharing infectious diseases data and samples	

The first round of coding was undertaken concurrently with data collection, facilitated by NVivo 12 qualitative data management software. The role of trust and perceptions of data and sample ownership were two notable emerging areas of discussion in the first set of interviews, and I proactively followed these up in subsequent interviews.

During the analysis I was particularly attentive to aspects of the discussion that related to the aim and objectives of the study. As the study has a focus on perspectives on ethical issues, I was sensitive to discussions about the morality of a particular situation or course of action, conceptions of right and wrong or good and bad practices, and the values that interviewees discussing as guiding what ought to be done. I was also attentive to discussion of dilemmas and difficulties in deciding what ought to be done, and contextual factors that affected the interviewee's perception of the ethical issues involved. I was open-minded in my approach to recognising discussions relevant to ethics by including those issues that were problem areas for people, even if they were not obviously ethical in nature on initial analysis. For example, one area of discussion for the location of data servers, which while not obviously 'ethical' on first reading, was clearly of concern to some interviewees due to worries about power and control. I therefore included this in the findings. Governance mechanisms were broadly taken to be formal procedures and processes used to enact rules and policies, with the focus of this study being on governance in relation to ethics. The interviewees were aware of the explicit focus on ethics and governance in this study.

Alongside regular monthly meetings with my supervisors during which we discussed the collected data, the codes and interpretations, I organised two qualitative analysis meetings with a wider group of researchers within Ethox. This provided the opportunity to examine selected anonymised transcripts in more detail to gather additional perspectives on the interpretation. I found these meetings very useful to consider alternative ways of understanding the interview data. I also presented drafts of the findings to both the Ethox research group, and to IDDO, as a form of sense-checking and to gather views on the presentation of the theme(s) and alternative perspectives.

3.5.2. Presentation of analysis in the thesis

Chapters 4 to 7 present the results of my analysis of the data collected through the semi-structured interviews. The data are not being used to directly make normative ethical claims, but to elucidate areas of practice within data and sample sharing in the infectious disease context that have ethical dimensions. I have not used pre-specified definitions of particular concepts that have normative significance but have aimed to remain faithful to the language used in the interview data. Quotes from the interview data are presented, and I have added emphasis (bold) to draw attention to key areas relevant to the analysis.

3.6. Ethical considerations and clearance

The study was approved by the University of Oxford OxTREC research ethics committee, as a minimal risk study (ref: 595-16).

I did a careful assessment of the possible risks and came to the view that the potential risks to interviewees related to this study were limited. I noted that interviews on the topic of infectious disease data and sample sharing may provoke politically sensitive discussion. As a result, there was an identified potential risk to the interviewee through disclosing politically sensitive information or views which could be linked to them. In order to mitigate this risk, interview data are kept confidential and all notes are securely stored from interviews and fieldwork, with any identifiable information only stored on the Nuffield Department of Population Health secure drive. All personal names and place names were removed from transcripts that were used in the analysis, and that were shared with others as part of quality assuring the data interpretation.

In addition, a further identified risk was that interviewees could perceive that the study may be used to criticise their practice. The aims and objectives of the study were discussed as a part of the

process of informed consent, and the study was not looking to highlight areas of bad practice in a particular initiative.

Participation in this study was voluntary, with no pressure to be interviewed, and the option to withdraw from the study. The interviewee could refuse to answer a question or terminate the interview at any point – although this did not occur for any of the 28 interviews, and all interviews were concluded once there were no additional questions on my part, and the interviewee had stated they had nothing to add, or when we had come to the end of the pre-specified amount of time the interviewee had given. Once I had prepared a near-final draft of the thesis, I emailed interviewees to ask if they would like to review use of their quotes. Four interviewees requested to review their quotes, and they subsequently did not request changes to the use of their quotes or the surrounding descriptive text.

3.7. Reflexivity: Reflections on my role as a researcher

The question for the DPhil arose from my own practice, as outlined in Chapter 1. Throughout my career to date as a public health doctor, and particularly when involved in responding to the Ebola outbreak in West Africa in 2014/15 as a health adviser at the Department for International Development, I have confronted questions around data and sample sharing. While this has given me great motivation and some background knowledge relevant to this research, I am aware that I have my own perspectives on data and sample sharing, including a sense that it has the potential to bring important public health benefits. As a result of my professional background I have also incorporated some public health and medical terminology into my own vocabulary and I am aware that I am not always conscious of moving into the use of 'jargon'. I worked to develop an interview guide, and style, that is attentive to a broad range of potential areas of discussion from interviewees with different professional backgrounds and have had regular discussions with my supervisors as the research has progressed. I organised coding meetings, as noted in section 3.5.1, in order to

gather a wider range of viewpoints on the interpretation the interview data, in addition to regular discussions with my supervisors on the research findings.

I am also aware that other characteristics – that I am a student based at the University of Oxford, and that I am from the UK and speak English as a first language - will have potentially impacted on interviewee perceptions of me, and the nature of discussions during the interview. In particular, the IDDO core team is also based at the University of Oxford. While I stated that I was based with Ethox, a research group in a different department of the university, this may have affected the way the interviewees from the IDDO network responded to my questions, if they considered me to be linked to IDDO.

In addition, I only conducted interviews in English – although sometimes it was not the interviewee's first language. I did approach a small number of non-English speakers for interview with a view to organising translation if required, but none of those approached agreed to participate.

Attending meetings related to data sharing brought its own particular challenge, which was to ensure my role as a researcher was well understood by those attending the meetings, who would at times draw me into the conversation as if I were an ethics adviser, as I had introduced myself as a doctoral student in a bioethics research group. I would be asked my views on whether a particular approach, for example in relation to consent (among other issues), was ethical. I found this to be a difficult interaction as firstly, I was there as a researcher not as an ethics adviser and I feel that these are distinct roles. I believed that giving my opinion would disturb the conversation and impact on the views of those present. Secondly, even if I had considered it appropriate to have given my opinion, I did not feel equipped to advise on the ethics of a particular approach, as I am a student in this topic at the early end of my journey learning about ethics, and am not trained in the way that, for example, a clinical ethics adviser may be when giving ethics input in relation to clinical cases. This issue in social science research was discussed by Charles Bosk, in his book

"What Would You Do?: Juggling Bioethics and Ethnography" where he reflected upon the difficult position of observing conversations as a researcher while avoiding being an active participant in them (135). Bosk was a Professor of Sociology, who had conducted ethnographic studies in healthcare settings. In this book, Bosk described his own "hesitation, doubt, and uncertainty" when asked his view on what to do in a particular situation (135 p.xvi), and outlined his own position – one that I have come to agree with through reflections over the course of this study – that over the period of data collection the researcher needs to take a step back and refrain from answering this question. He discussed a number of reasons, such as not having the required knowledge, and impacting upon the object of study - becoming entangled within it, to a greater degree than desirable when doing research. In keeping with this position, I was careful to outline my role as a researcher studying perceptions on the ethics of data and sample sharing. I found it uncomfortable to have to reinforce this and feel that I was being unhelpful, re-stating that I was not in a position to advise on the ethics of data and sample sharing during the meeting, but I hoped my research would be helpful on reflecting on the discussed issues in due course. I think I found it particularly uncomfortable as prior to undertaking the doctorate I was working as a public health professional actively providing input into such meetings and discussions (from a public health perspective).

Another related issue is that my presence in these meetings may have altered the nature of the discussions. It is notable that ethics was frequently and explicitly discussed in relation to the data sharing platform, and this may have been because those attending the meeting were aware that someone researching ethics was present in the room. It is impossible to say whether the nature of the discussion during the meeting would have been different if I were not in the room, but I found it notable that it was such as prominent topic. I took notes during the meetings for my own records, summarising the nature of the discussion without attribution to individuals, and I have kept some reflective notes on these interactions.

3.8. Conclusion

Following the identification of a research gap in Chapter 2, this chapter has provided a description of the chosen methods for this study and the rationale for these methods. The methods evolved beyond an examination of data and sample sharing in relation to specific diseases, to a wider study including interviewees with experience across a range of infectious diseases, based in different geographical areas – predominantly in sub-Saharan Africa and Europe. The next four chapters will present my analysis of the interview data, and the presentation of major identified themes.

4. Typology of data and samples, and the ethical implications of sharing them

This chapter presents a typology of data and samples being shared in the infectious disease context. The type of data and samples were perceived by interviewees to have distinct implications for considerations relating to sharing. This was an important theme emerging in my analysis of the interviews. The first section of this chapter (4.1) focuses on the different types of data and samples relevant to infectious disease related sharing. The second section (4.2) outlines the different ways in which sharing takes place, and how the type of data and samples impact upon what is considered morally permissible with respect to sharing them. The third section (4.3) shows how different facets of the infectious disease context influence what is perceived to be morally permissible with respect to sharing data and samples. The intention in this chapter is not to provide a comprehensive typology of all possible types of data, samples, infectious diseases and sharing activities, but to set out a typology based on the practical experience of the interviewees.

4.1. Typology: Data and sample collection

A number of types of data and samples were referred to by interviewees in this study in relation to infectious disease research and responses. Interviewees in this study generally focused on the data or sample types most relevant to their own work. Some interviewees worked with a variety of types of data and/or samples, and made comparisons of the issues between types of data or samples.

4.1.1. Differences between data and samples

4.1.1.1. Samples as a form of data

Interviewees saw data and samples as connected or distinct to different degrees. Some interviewees perceived a clear interconnectedness between data and samples, with one (interview 28), a researcher based in Africa, stating that "[d]ata can exist on paper, can exist electronically, can exist in the form of samples or tissues, for that matter. So anything from which information can be extracted is data." Data was seen by this interviewee as an overarching term encompassing samples.

4.1.1.2. Samples distinct from data

Although some saw data and samples as closely connected, the prevailing view among most interviewees was that samples had certain features that set them apart from data. Firstly, the finite nature of samples was contrasted with data, which were viewed as replicable. This illustrated by the quote below, in which the interviewee explains that this makes data more widely usable by a number of people simultaneously:

> *Data, once you have it, isn't a limited resource. I mean, you can give it to as many people as you'd like. Samples are [a limited resource].*
>
> **Interview 12, Doctor and Laboratory Director, European, Government Agency, Europe**

A further difference between data and samples was the tools required to use them. The interviewee quoted below suggested the need for specialised equipment to use samples, compared to data which could be analysed with more everyday equipment:

> *…I am not [an] epidemiologist or a researcher, but there's probably something I can do with data, but I cannot do anything with samples, because if you do something with samples, I need to have a lab. I don't have a lab, I need to be a lab technician, I need materials. With data I just need a computer. Even if I don't have a computer, I can still do it probably with a calculator, that's how people used to do it before.*
>
> **Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe**

An additional area of difference highlighted in interviews was that samples were perceived by some to be more closely linked to the human body (in the case of human samples). This discussed in the following quote, relating it to superstitions people may have around samples:

> *…you know, some people are superstitious about blood. Or a lot of superstitions, bloods and samples seem to be the locus of a lot of superstition. It's hard to be superstitious about data. It would be very abstract.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

The interviewee used the word "abstract" which highlights the distant and removed nature of data.

4.1.1.3. Data and samples: Differing perceptions

While many considered data and samples distinct in particular ways, the perceived importance of these features varied. The interviewee quoted below, a doctor based in a European non-governmental organisation (NGO), explained that in her view it is more important to consider the subject matter that the data and samples relate to, and that individuals will make different judgements about whether data or samples feel more personal. In the case of samples, she related this to differing social and cultural meanings, suggesting that Africans may have particular beliefs about samples being a part of their soul. This links to the discussion above, and the idea that samples remain linked to the person from whom the sample was obtained. Conversely, she noted that data about a sensitive subject may be considered to be more personal than samples to other individuals:

> *First difference [between data and sample] is perception.* ***Perception, because for a lot of people, having data is not the same thing as a sample. And you can have the exact opposite perception.*** *For example,* ***one of the perception could be, from let's say Africans, that a sample is a part of them that has a part of their souls,*** *and so you can't just do whatever you want. Others would say ... You're capable of getting very specific identifier information from me, because you can have my genetic material out of it. These are sort of the ones who say that it's much more tricky to have samples, as compared to data.* ***Others would say that it depends on the data.*** *If the data would be about my blood pressure and my weight, and my whatever, cholesterol, I'm just inventing a bit. Then they really couldn't care less, because that's not really very important, because all these variables are changeable. I can lose weight, I can gain weight, my cholesterol might be different, and certainly my blood pressure. Or you have data that includes much more personal data, that would include the number of*

abortions I had, or something. The fact that I was raped, with a bit of details. Or issues related for example, mental health. Mental health issues. These kind of data they define who I am, much more than my cholesterol level or my weight. ***Therefore, they would say that getting my genetic material, my DNA, who cares, but getting my story, my story is who I am. You see? So, you have different perception about what it means, data.***

Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe

This is particularly relevant for infectious disease responses, where data and samples may be shared rapidly across different places, where sensitivities vary. This interviewee also discussed the stigmatising nature of certain information that the data may reveal. Perceptions around data and samples related to a stigmatising diseases will be discussed in more detail in section 4.3.2.

4.1.2. Types of data and samples

A clear theme arising in this study is how interviewees tended to categorise different types of samples and data in different ways. These categories seem to have moral relevance, relating to differing sensitivities of the data and samples types. The types of data and samples fall into two broad categories – firstly, the type of data and samples that are collected, and secondly, the type of data and samples that are stored, as illustrated in Figure 4.1. In what follows I present this typology, that emerged through the interview data analysis.

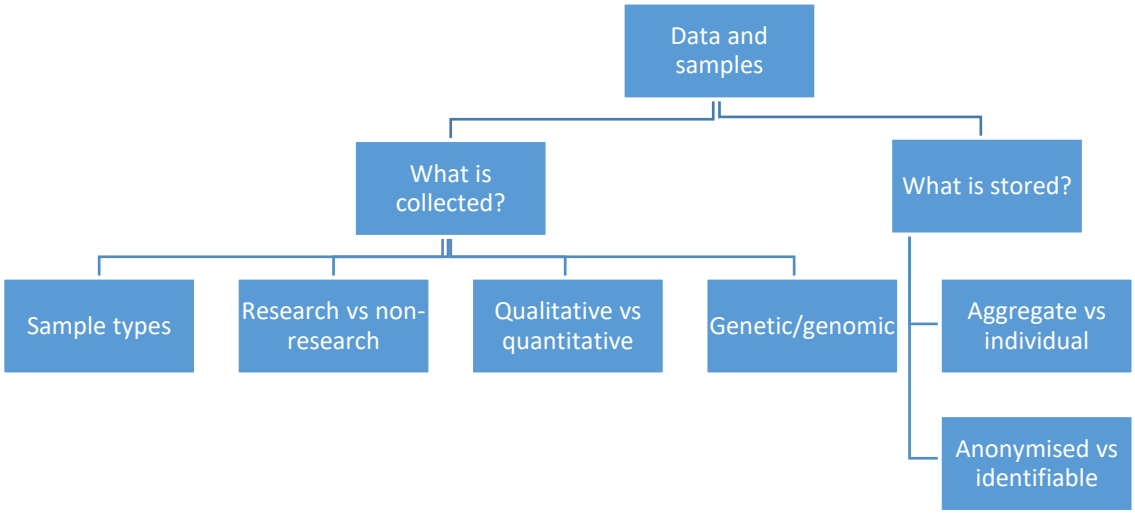


Figure 4.1: Diagram of data and sample types from interview analysis

4.1.2.1. Different types of samples

A range of sample types were discussed by interviewees. The term 'sample' incorporated a heterogeneous set of things, as illustrated by this quote discussing the development of a biobank of samples collected during the Ebola epidemic in West Africa:

> *Basically they [the samples collected during the Ebola epidemic] were all research samples - for example, they were primarily blood. But we also had saliva, we had tissues, we had semen, breast milk. We had ... vaginal secretions. We had…mouth swab from deceased participants. We had plasma. There was a wide range of samples.*
>
> **Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa**

These sample types were viewed differently. Particularly noted in this study were the sensitivities associated with blood, and with embryonic/foetal tissue, which may have particular significance or beliefs associated with them. Echoing the discussion above regarding samples, the following two interviewees discussed the sensitivity of blood, particularly in parts of Africa where it may be seen as a part of that person, or have other uses such as in witchcraft:

> *There are places where blood is very sensitive, it's even used in witchcraft and things like that, or someone's placenta or someone's urine.*
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**
>
> *Certainly the case in Africa, blood is very important. People think that, you know, if somebody takes your blood, that's like taking a piece of you. It is, but you know, I wouldn't consider that. If I gave blood and somebody walked off with it, I would immediately sort of divest myself of any concern for it.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

A small number of interviewees also discussed the use of embryonic/foetal tissue as being both useful for research but also particularly sensitive. The interviewee below discussed an issue that

arose in her work, where there were discussions about using foetal tissue from a patient with an infectious disease for further research, because of its particular utility:

> *My understanding is that an unborn child is potentially of very high value for research purposes when it comes to disease. So we talked more about blood samples and potentially finding some kind of vaccine or cure for Ebola, whichever everyone would like to see. An unborn child obviously provides ... well not obviously to me but my impression was that for medical practitioners who understand medical research, that an unborn child is a sort of, you know, high-value research product or whatever the wording is. It was very driven by that.*
>
> **Interview 3, Policy Maker, European, Government Funding Agency, Asia**

She had also said that in relation to research on such tissues, "we all agreed that that shouldn't happen because it wasn't ethical" – linking this to the particular sensitivity of such tissues, and the need for consent for any research using them. This shows that while this interviewee was aware of the potential high value of such samples to understand the disease, the nature of the sample was relevant to what it was permissible to do, and this was discussed explicitly in terms of ethics.

Most interviewees focused on human samples, but a few also discussed pathogen samples, some of which may be derived from human samples. The additional biosafety requirements for certain types of pathogens and their potential to be used in the development of bioweapons was noted. The high value given to certain pathogen samples, in this case relating to Ebola, because of their potential to be turned into biological weapons is discussed in this quote:

> *Ebola and their samples fall within the purview or the interests of the global health security agenda, because clearly, it is recognised that these could, by extension, be developed as some kind of biological weapon if people had an interest in expanding the use of the pathogen in that direction. So, there's so much interest in these kinds of samples, and anyone who has them has a valuable biological commodity.*
>
> **Interview 28, Doctor and Scientific Director, African, Research Institution, Africa**

4.1.2.2. Research and non-research data and samples

A further way interviewees categorised data and samples was between those that were related to research and those that were not, although there was not consensus about the practical implications of these distinctions. Non-research data and samples were those collected through clinical care and public health activities such as surveillance. Research data and samples were generally discussed as being collected through clinical trials, epidemiological and social scientific studies. The quote below illustrates the distinction being made between research and non-research data and samples:

> *…there are sort of two or perhaps three sources of data from the Ebola treatment centre. One is routine clinical patient data, data that were collected in the process of admitting, caring for, and discharging patients. There is another body of evidence which relates to the samples that are collected. Either samples that were collected routinely for clinical care, or samples that were collected ostensibly for research purposes. There were also data collected, paper data and patient data, collected for research purposes, explicitly for research.* ***So there are kind of three sources there, and there are different considerations depending on what we're talking about.***
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

The interviewee notes at the end of the quote that "there are different considerations" – this relates to the sharing research and non-research data and samples, which will be explored more in section 4.2.3.2.

4.1.2.3. Qualitative and quantitative data

While interviewees were mostly focused on quantitative data, or on samples, two interviewees did discuss the possible differences between sharing of qualitative and quantitative data. They discussed the nature of collecting and understanding qualitative data, stating that the data reveal particularly personal details about the person being interviewed, are difficult to anonymise while retaining meaning, and are contextual - collected through an interaction with the researchers that others may not understand.

The degree of difference between qualitative and quantitative data was, however, questioned by both interviewees, who suggested that quantitative data may also be very personal:

> ***Qualitative data can get a lot more personal.*** *Someone will share something with a qualitative interviewer that they wouldn't talk about otherwise. But that's not quite right.* ***Quantitative data can be really personal as well.*** *People can share stuff there that they wouldn't otherwise, even if it's just the answer is like three, or something. It still might be something you would never tell anyone, so that's not quite it. That's not the difference.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

The relevant distinction for the interviewee above did not seem to be whether the data are qualitative or quantitative in nature, but whether they relate to information the person that the data were collected from would not usually divulge, and may be doing so within the trusted relationship that has developed with the specific researcher undertaking the data collection.

4.1.2.4. Genetic/genomic data and samples

Genetic or genomic data related to both the pathogen and host may also be collected and shared for infectious disease related research and responses. This may be to understand genetic susceptibility to the disease, or to track super-spreading events, for example (136, 137). Data and samples that relate to human genetics were noted by a small number of interviewees as being particularly sensitive. The following interviewee reflected on her own uncertainty as to why this is but suggested that it may be because the genetics of an individual also provide information about those to whom they are related. She went on to consider how this may "expose" the individual's village or country:

> *I really don't know because I don't know much about genetics but there's talk about I represent an entire village and that village represents an entire county. And so you don't know what people can do with your genetics. I really don't know what they would do with them but I hear that kind of talk. Just getting genetic mapping and everybody can, "We know everything about you," or something like that. I don't know…* ***I think it's that thing of exposing your entire village and your entire county just through this one person's genes.***
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**

One interviewee, a Scientific Director based in Europe, also considered that genomic data presented more challenges with respect to differing capacities of low- and middle-income countries (LMIC) and high-income countries (HIC) to analyse and use:

> ***I think we need to acknowledge that genomic data are still a toy of rich country which cost a lot of money with very few individuals who have the capacity to analyse data.*** *The translation of capacity into affected countries is even more complex than data. I mean data, individual pooled analyses, are complex and you need to find a good statistician. Genomic data, well you need to find a good statistician used to genomic data and those are more rare at this stage.*

The above quotes, taken together, show genetic and genomic data were perceived to present more complexities, compared to other types of data, both with respect to what they may reveal, but also practicalities about who has the capacity to analyse and use the data.

4.1.2.5. Aggregate and individual level data

One clear line of distinction that emerged in the analysis was between aggregate and individual level data. A number of types of data were discussed in the interviews at the individual level, relevant to infectious disease research and responses. These included:

- Clinical trial data, gathered in the process of conducting experimental trials. In this study the most discussed clinical trials data were those relating to Visceral Leishmaniasis (VL);
- Clinical data, gathered in the process of clinical care;
- Survey data, with the Demographic and Health Surveys[10] discussed by some interviewees in this study;
- Notification, line listing and contact tracing data, gathered through notification of people with an infection of public health interest, and their contacts;

[10] The Demographic and Health Surveys are surveys conducted in low and middle income settings, aiming to collect population and health-related data. This can include data related to infectious diseases – for example, immunisation data, and data related to HIV/AIDS, and malaria. For more information see: https://dhsprogram.com/ (last accessed 1 October 2020)

- Genetic and genomic data related to both humans and the pathogen;
- Case report data, in this study discussed in relation to reports of cases where an unusual or new treatment was used to treat an infectious disease that does not have many associated treatments;
- Qualitative data, discussed in this study as data gathered through social scientific study.

Aggregate data are data that are compiled, so that they no longer relate to individuals. Aggregate surveillance data gathered during an outbreak was compared to a sample poll of people walking past a location in this quote:

> *However, if you were doing a sample poll on [anonymised], the number of people that are walking past the Tescos, you collect my data but you don't ask me for consent. I think it's still research, but I am a number, I am not ... my individual information isn't being used, I'm being used as a number without my consent. To a certain extent, you have the same ...* ***in terms of number of cases that are being counted during an epidemic. You wouldn't have someone's case to say I'm counting you as a number in this epidemic. That's data. It's anonymised data. So I think a lot of data is anonymised, and actually for someone to be used as a number is no consent required.*** *If you're using their personal information as part of a study, for example, and some of the information they might have provided in their patient history, that's important and consent is required.*
>
> **Interview 14, Doctor and Policy Maker, European, Government Funding Agency, Europe**

Here, the interviewee outlined a contrasting feature of aggregate surveillance data as compared with individual level data – that it does not provide information at the individual level, and this makes it a form of anonymised data. He also discussed the resultant implications for consent – suggesting that given the lack of individual level data it is not required for this form of data. The category into which data fall were therefore seen as having implications for their moral status and the kinds of regulatory permissions needed to use them.

4.1.2.6. Anonymised, de-identified and identifiable individual level data and samples

In relation to individual level data, a further commonly made distinction made by interviewees was between anonymised and identifiable data and samples. The possibility to anonymise individual level data was discussed, where the data are altered in such a way that the individual from whom the data or sample had been collected cannot be identified. Anonymisation was seen as a way to protect participants from risks that may come from being identified with an infectious disease. De-identification, where individual identifiers are removed (138), was also sometimes discussed although the distinction between this and anonymisation was not always made clearly by interviewees.

Some felt, however, that it was not always appropriate to anonymise data, and that there were certain instances in which individuals needed to be identified to enable research, as illustrated by the quote below:

> *You can probably think of examples where actually having anonymity is counterproductive but that's something you have to deal with on an individual basis. So that is anonymity on where the data came from, but then it's equally important that if data are to be of any use that the essential information is linked to it. The information about the presentation, age, sex, pre-existing medical conditions.*
>
> **Interview 12, Doctor and Laboratory Director, European, Government Agency, Europe**

In summary, while identifiable data and samples were often seen as riskier, with more potential for harm, than anonymised data and samples, anonymisation was seen by some as potentially reducing the utility of the data and samples.

4.2. Sharing of data and samples

Much of the discussion about the different types of data and samples related to the implications for their movement – i.e. sharing them. In this section I start off by outlining the forms of data and sample sharing that were discussed by interviewees, before then going on to explore

perspectives on the interactions between these different modes of sharing and the different types of data and samples set out above.

Within the interviews the term sharing was used in different ways by different people. Data and sample sharing sometimes involved sharing with known others within the scope of teams and collaborations, and sometimes took place through data repositories and biobanks with the implication that those who had access were mostly unknown to those who were responsible for the original study. It was sharing with unknown others that more often seemed raise potential concerns for those who were data or sample contributors. A further interpretation of sharing given by some interviewees was that of clinical results with healthcare workers and patients, sharing blood specimens from patients with a laboratory, or wider sharing of clinical case information. This suggests that sharing is a term with different uses, and there are a number of ways in which people perceive that they are sharing. However data and samples are shared, the wider information required for sharing to be effective was discussed by interviewees, such as the sharing of metadata and data dictionaries to aid their use and interpretation.

4.2.1. Data and sample sharing with known others

One type of data and sample sharing is the type of sharing that would form a part of someone's day to day work - sharing with their boss or within their team. The utility of such data sharing, which was seen as a necessary part of work, to harness the skills of team members and check analyses, is illustrated in the following quote:

> ***Sharing data, there's one sense of sharing data, which is sharing within a team.*** *So you'll have a team of people who are working on a study, and there's quite a strong sense, I think, of collaboration within that team. That everyone should have access to the data within that team. I mean it's not sensible to have just one person analyse the data and tell everyone what the result is. You know, in the very…most basic sense you need somebody else to be able to check it and verify it, which means other people have to have access to that data. Just the other day I shared a dataset with my boss and I sent her the analysis files so that she can check it.* ***But that's sharing data, but sharing with someone I work with all the time. And it's sharing anonymised data in secure conditions with just one other person. It's almost the minimal level of sharing that you can have.***

Interview 2, Epidemiologist Researcher, European, Research Institution, Europe

The interviewee described this as a minimal level of sharing, which would not generally be considered as such according to the guidance discussed in Chapter 2 which tend to assume that sharing is outside of the original core team collecting and working with the data and samples.

Sharing within a wider collaboration was also discussed. The wider collaboration may be geographically dispersed, or based in different institutions, but the sharing here is still with people who are known to the people contributing the data or samples. The rationale mirrors that of sharing within teams, enabling researchers to harness the skills of others. It was also seen as an opportunity to learn from others and build capacity:

> *I think in my view, there's a lot that people could share, or learn from each other, even in terms of collaboration. Even skill sets, one can learn in terms of hunting the data. Yeah, skill set or handling data. And in a way, sharing experiences. I think in a way, that can lead to capacity building, in terms of strengthening ... Like we adjust skill sets and maybe building new ones.*
>
> **Interview 8, Clinical Trials Manager, African, Research Institution, Africa**

The interviewee went on to suggest that consortia or collaborations are a collective endeavour that allow people to feel included. She discussed the sharing of experiences through sharing the data.

This type of sharing, with known collaborators, was contrasted with sharing that involved unknown others and suggests a sense of being either insiders with whom people share with more easily, or outsiders with whom there may be resistance to share:

> *Well at least a lot of reluctance and not really big issues but at least really yearlong discussions whether for example data should be made open access or should be shared in a broader sense than just within the consortium. Within the consortium, I could do any data request and I will get it without anybody asking anything almost as long as I, of course, get a justification for it but I'm not sure why there's such a difference between within and outside the consortium.*
>
> **Interview 4, Pharmacologist Researcher, European, Research Institution, Europe**

Although sharing with people within a team or collaboration was generally perceived to be more straightforward by the interviewees in this study, a risk of this type of sharing was discussed by one interviewee. She had experienced identifiable data being shared with her unnecessarily as she was known to the person doing the sharing and felt that as a result, the person doing the sharing was happy to divulge this information. She went on to outline her view, that the protections around data should be the same, as if it is being shared with someone who is not known to the person sharing the data:

> ***You get sharing data within a team and it can be a problem because sometimes people want to share non-anonymised data.*** *You get the people working on the data in the field, they know the participants because they actually see them. Sometimes they're not as strict about confidentiality as they ought to be, because they know who all the people are anyway. So even if you take the names off and you just use the ID numbers they still know who those people are, and I think sometimes it's easy to forget that other people don't. So then they'll share data with someone like me, and it's still got names on it and I write back and say, "Please don't do that." And they say…I got a case of this recently, and she wrote back and said, "I thought it was okay because you're part of the team."*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

4.2.2. Data and sample sharing with potentially unknown others

Sharing data and samples with people who may not be known to contributors, through data repositories and biobanks, was also discussed by interviewees in this study. Both open and managed access repositories were discussed. As noted in Chapter 1, following the Human Genome Project many funders initially called for full and rapid open access to data. There have now been increasing moves for managed access. Sample biobanks were discussed by some interviewees, to collect together samples to facilitate research.

Data sharing into a repository was seen as something beyond the day to day sharing that goes on between team members or collaborators. By enabling wider access, it would enable people with a wider range of skills to use the data or samples, and people to approach them with questions that those collecting the data or samples may not have thought to research:

> *We put things in repositories. We don't know where they'll go. All such things could happen. They'll start being combined together. We'll get people grouping together different data sets from all over the world and coming up with huge ones. And yeah…Or combining it with ... You know,* ***looking at it from different perspectives.*** *You get economists looking at health data. You get psychologists looking at education data. You get people coming in with a whole different mind-set looking at it in a completely different way. That ought to be a positive thing.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

The quote above shows that value that such data sharing is perceived to have, as it can open up new avenues of exploration by people from different disciplines.

The purpose of the data repository was usually to allow pooling of data relating to a particular disease, in order to increase sample sizes and allow meta-analyses to be conducted. This was discussed by the following interviewee, who explained the possibility of revealing "hidden knowledge" through these larger sample sizes:

> *So data sharing in my understanding is pooling data together from different research groups, from different clinical trials in different countries, of course in related clinical trial areas, to be able to be used for data analysis. You know, you pool data together, you bring in to one big pool, and you get better statistical power to be able analyse this data together. So data sharing is all about different clinical trial institutions, organisations, investigators. Bringing respective data together in to one pool.* ***This data can then be used in meta-analysis to be able to understand more about the disease areas, and to be able to be reveal hidden knowledge that cannot easily be seen in clinical trials****, let's say with the sample size of 50, or less than 100.*
>
> **Interview 19, Data Manager, African, Research Institution, Africa**

One interviewee discussed developing a repository for data sharing as a scientific endeavour in itself, explaining that data sharing was an active process, requiring careful compiling and harmonisation of the data in order to render it useful for science. This was seen as an advantage to this form of sharing – that the data are compiled by platform conveners in a considered way that enables them to be useful for important scientific or public health questions:

> *…you can't just take a bunch of data and throw it in a Dropbox and say, "Oh we've got our data and it's shared, it's great". Actually, there's a tonne of science that needs to go into the design and framework of the technical architecture of the data so that people can understand it. So that* ***it can be properly***

harmonised amongst different datasets and so that it's designed in a way that can deliver science.

Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe

However, some interviewees were uncomfortable sharing with a wider group of people through a repository, as discussed in the quote below. The interviewee argues that the term 'sharing' was a problematic term for repositories. She suggested that the term sharing is morally loaded, and an inappropriate descriptor of the activity being undertaken:

Well, I think, and I think…thing is that it's not actually sharing. It's data giving in most senses. And one would expect that sharing is to a certain degree reciprocal, and ***a lot of that…initiative is about just giving your data to them****. And insofar as sharing, or a reciprocal arrangement is created, I don't see a mechanism for that. I don't necessarily think it should be called data sharing. It should be called data giving, and that's fine. I'm not opposed to data pooling, or having a third party manage your data. But* ***I would also, like…kind of contest this word sharing.***

Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe

The reference to "data giving" suggests that the interviewee did not feel that there is inclusion in the process, rather, the data have been handed over to someone else

4.2.2.1. Open access data

The nature of sharing beyond collaborations can be through a repository that renders data open to a range of people, or manages access so it is closed to specific people.

The increasing expectation within the scientific community for more widely open and accessible data was highlighted by a number of interviewees, particularly during infectious disease responses. One interviewee recounted the changes he had witnessed over time, with an increasing expectation to make it more widely accessible:

So it could be, even from about 10 years ago, then there was one project we had to write the data sharing statement. And more or less what we said was we'd share it in such a format, more or less to bona fide

> *researchers essentially. So that's the kind of minimal, you know … In a way that's the kind of minimal data sharing.* ***Whereas the, what we're coming onto nowadays is more like you have to pretty much ... Well, under some circumstances you have to pretty much like make the whole dataset available at the individual level, subject to attempting to de-identify it.*** *Meaning, to put it on some website where really quite a wide range of people could access it, up to it being even quite openly accessible.*
>
> **Interview 20, Medical Statistician and Epidemiologist, European, Research Institution, South America**

The interviewee also used the term "*attempting* to de-identify it" (emphasis mine) suggesting a scepticism that datasets can always be fully de-identified.

There was a prevailing sense of scepticism and concern around open access data among interviewees with a range of backgrounds. Alongside concerns about the risks to research participants or patient (discussed further below) interviewees questioned whether data could truly be considered open. For example, when discussing open access, a small number of interviewees noted that they may not be open in a meaningful way. One reason given was that some data, such as genomic data, may be specialised and therefore only really usable by those who have the specialist knowledge and not accessible to others more widely:

> *But anyway we're trying to think of other models and because I'm in the department of genome sciences I've watched my colleagues for a very long time have these huge studies where they got more than I'm even naming here and they've developed a model where all of the data were open but they're really not very accessible when you ask them about it? Well, how am I going to use this information? What's open is the raw reads, which if you're part of the in-group you know how to use that information but if you're not then and I'm sort of not really then I can't really use that information without ...* ***It's a little disingenuous to say that their data are open, I mean it's true but it's not really open and accessible.***
>
> **Interview 16, Scientific Director, North American, Research Institution, North America**

Another reason given was the need to keep data confidential and remove identifiers when making them open access, the data are rendered meaningless, and so a controlled structure for access with a fuller amount of data would be required:

> *I think it's very, very rare where you literally have open access. That open access's got to be data, which has no identifiers whatsoever really. And* ***without identifiers, data are pretty meaningless without some sort of other data.*** *So I think, open access is always within some controlled structure.*
>
> **Interview 15, Doctor and Scientific Director, European, Research Institution, Africa**

Another concern raised about open data was that, depending on how it is done, it may potentially be more difficult to use or harder to find the data, as they may just be posted somewhere on the internet without proper collation or signposting:

> ***It's not just about putting something in the public domain, because actually things can be easily hidden in the public domain****, you put out so much data no one knows where to find it. It's about pro-active targeting of data and evidence about the people that most need to use it. It should be nuanced, it should be targeting the people that need to know that specific piece of evidence, that specific piece of data.*
>
> **Interview 14, Doctor and Policy Maker, European, Government Funding Agency, Europe**

This quote also raises another point – that open data was sometimes seen as synonymous with uploading it without much thought onto the internet, although in practice a carefully curated repository could still have an open access policy.

All of the interviewees in this study who discussed open data noted some risks or issues with completely openly available data that had no restrictions on who can access it, and the purpose of access. Various concerns were raised, such as the need to keep some data private, and managing underlying issues with the quality of the underlying data. A further critique was that the data may be used for 'bad' research. The interviewee below critiqued the idea that proper data sharing is completely open access, for anyone to use for any purpose. She suggested this risked poor quality research being done, and that this was not what data sharing was about:

> *You see, the broader public would think that data sharing is that you put everything somewhere and people can grasp whatever they want, and do whatever they want with it, and do whatever bad research with bad*

protocols and bad methodology and come up with bad results. This is not data sharing, right? This is not it.

Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe

Most interviewees, as noted above, felt that sharing data in a way that was fully open access without managing the access in some way was not desirable. One interviewee however did perceive a benefit. She noted that making data open, without a managed access procedure would reduce the time that individuals needed to spend reviewing proposals in order to decide whether access would be permitted. She discussed the journal peer-review system as another route for quality assurance:

The advantage would be that people don't need to review new proposals and spend time. *So it's … Because I think at the beginning they will be highly motivated to do that, but little by little they will be tired by receiving ... Maybe not hundreds, at least for VL, probably it would be more smaller than that. But it's just that it takes time and I don't think that people will have this time in the future, on the long term.* ***The disadvantage is that you may have some bad quality analysis that could be done, because if there is no control about the scientific communities around, it could be that some results are biased and not well conducted, so this might create some kind of noise in the scientific community.*** *I mean background noise, I mean. Of course, the publication would be peer reviewed, but you never know exactly how it goes.*

Interview 7, Clinical Trials Manager, European, Research Institution, Europe

While the interviewee discussed the role of peer-review, she also shows some scepticism that it would be a sufficient safeguard against poor quality analyses.

4.2.2.2. Managed access repository

As a counter to fully open access data, one commonly discussed way of sharing was that of managed access data repositories. Some of the interviewees were selected because of their involvement in developing data sharing repositories, so this may have framed their interpretation of sharing. All sample sharing was discussed either within a collaboration with known others or a managed access repository (biobank), and no interviewees raised the possibility of open access to samples.

Controls on who can access the data, and for what purpose, were felt to be important by a number of interviewees, particularly by those who would be contributing to, or were developing, a sharing platform. Interviewees discussed these controls as a way of allowing more vetting of who is using the data and samples, and the purpose for which it is being used. Contrasting with the discussion in the previous section on open data, interviewees saw managed access repositories as a way to prevent poor quality science being conducted. This was explained by the following interviewee, who noted that without such controls poor quality science could devalue data sharing:

> *…it's great to grant access to data, but* ***I think it would be important to determine who has the capacity or what is the minimum requirement that you can demonstrate that you have to analyse complex set of data.*** *Because the risk is having people who are unequipped to do this kind of complex analysis and do wrong analysis and then we need to spend a lot of time into…re-challenging a secondary analysis or demonstrating that they did it wrong.* ***It's time that often people don't have to do this kind of effort and this will be actually begin to jeopardise the value of data sharing****.*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

The restrictive nature of these data sharing repositories and biobanks requires processes to determine who can access the data and samples and for what purposes – raising ethical questions about what constitutes a good purpose, and who ought to decide. A number of processes and oversight mechanisms were discussed by interviewees, such as data sharing and material transfer agreements, terms of submission documents, data/sample access committees, scientific advisory committees, and research ethics committees and institutional review boards. These processes were also seen as a way of providing an opportunity to set-up a prior agreement on issues such as intellectual property and the distribution of any goods arising from research on shared data or samples. The nature of these agreements, and their perceived role in good practice will be presented in more detail in Chapter 7.

4.2.3. Sharing different types of data and samples

The previous sections have explored the different types of data and samples used in infectious disease research and responses, and the different ways in which they may be shared. This section now explores these two things together: the implications of the data and sample types on considerations for sharing them.

4.2.3.1. Sharing samples compared to sharing data

As noted in section 4.1.1, interviewees made distinctions between types of samples and data. My analysis found these distinctions were linked by interviewees to what is permissible and impermissible to do with the data and samples.

One notable distinction emerging in the analysis was the replicability of data and the finite nature of samples. This theme is illustrated by the quote below in which the sharing of data and samples are contrasted. The quote also emphasises the issue of prioritisation – given the finite nature of samples, the need to carefully judge how best the limited resource is used.

> ***Well there's a basic difference that you can use data for ever, as much as you want and turn it upside down ad nauseum. For sample sharing…the use of the sample would have a limited spectrum.*** *And so, how do prioritise the use of those samples would be the key question, number one. Number two,* ***when you are talking about sample sharing, I suspect that there's a lot more intellectual property rights that may arise*** *or questions that may arise around intellectual property rights, which is complex when we especially talk about a market which is not driven by commercial interest. And so how do we in that case protect the patient in making sure that they will access whatever product will be generated from this intellectual property rights.*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

The outputs of research on samples were also considered to be potentially distinct in the quote above, with the suggestion that there could be additional intellectual property considerations on any products arising from research on samples.

As noted previously, samples were also considered to be more closely connected to the human body by a number of interviewees, and were generally discussed as more personal. The following interviewee both from and based in sub-Saharan Africa uses an evocative analogy likening samples to the remains of a body after death, and the feeling people have about what happens to their remains, and their desire for them to be located in a specific place. He used this analogy to suggest that samples should remain in the location that they are from – in this case relating to samples from sub-Saharan Africa – and oppose the idea of wider exportation and sharing of samples:

> *But I think when it comes to the samples of maybe blood, samples of tissue and others, I think that in line with our basic human ethic, these should be in Africa. I think all over the world,* ***you hear people wanting their mortal remains to be repatriated to their countries of origin. People like to have their ashes scattered in particular places, and so on and so forth.*** *So while I recognise that, yes, you may be dead and gone, or may be just a sample, this recognition of it being a part of a human being, no matter what it is, you want it to be located where it came from, especially if work is going to be done on it.*
>
> **Interview 28, Doctor and Scientific Director, African, Research Institution, Africa**

Data generated from samples, for example diagnostic work on clinical samples contributing to surveillance, or research data generated through work on samples was not discussed in a similar personal way as the samples themselves.

A further reason sample sharing was considered to be distinct was the potential to do unforeseen analyses on samples in the future, particularly with emerging technologies:

> *Completely different because the data you share, I mean everyone is aware and you cannot look at something that was not planned because the way we are gathering the data is very well defined. Whereas for biological samples, you don't know what people could do on DNA and so on. Usually they don't have much information to cross link, so maybe it's a fear which is not justified but in people minds, it's I think very different.*
>
> **Interview 7, Clinical Trials Manager, European, Research Institution, Europe**

These distinctions noted above lead to a number of additional questions with an ethical dimension for sample sharing, compared to data sharing – how best to prioritise the use of samples, what

future analyses should be permitted, and how to manage the personal connectedness. It is not, however, the case that data sharing was always perceived by interviewees as ethically more straightforward than sample sharing. The possibility of replicating data was noted to lead to a challenge to control and protect them. This interviewee explained the different storage implications of data and samples, explaining that samples require effort to move from place to place, enabling their movement to be more controlled. On the other hand data, however important they are considered, can easily be sent on to others, including unintentionally:

> ***It's hard to email tissue. You know, they have to be kept in the same place under certain circumstances.*** *So I think to a certain degree you have more control over them, but maybe also paradoxically less control. They need to be handled differently, and I think the basis on which I differentiate is that the arrangements made for them are different than the arrangements made for the data.* ***The data, I mean, the whole system of data management and data protection is just so porous.*** *I mean, you can have data and tell everybody, "This is the most important data in the world," you know? And somebody could just email it to somebody else by accident. You could say to somebody, "This is the most important tissue sample in the world," and they still have to go through an awful lot of logistical arrangements to move it from place to place, or get it out of a locked lab, or what have you.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

4.2.3.2. Research and non-research data and samples

A further distinction that emerged in my analysis was between research and non-research related data and samples, as discussed in section 4.1.2.2. Sharing these types of data and samples was contrasted by interviewees, and distinctions were made around how much information is contained in the data, and the importance of sharing the data and samples rapidly.

It is notable that the type of data and samples did not necessarily correspond to the type of use once shared, so the process of sharing altered the typology of the data and samples. For example, research data may be shared to facilitate the infectious disease response, and non-research data and samples such as surveillance data, or data collected through clinical care may be used for further research.

The interviewee below discussed the collection of routine data in epidemics. He considered that non-research data such as clinical data collected during epidemics would be likely to be less complete than research trial data, highlighting a perception that non-research data are more straightforward to share:

> *When you have an epidemic, you'll probably be just doing routine care, what is basic to get the patient in a better position. And in such a situation, we will probably not be collecting as much information as you would in a typical clinical trial setting. And the fact that you are not collecting as much information, I don't think it's a problem or it's wrong to still go ahead and share that data.*
>
> **Interview 13, Statistician and Group Head, African, Research Institution, Africa**

Although the quote discussed this in terms of whether data are collected through research or non-research activity, the distinction being made by the interviewee (as with the section on qualitative data) seems to be the amount of personal information in the dataset.

Some viewed sharing data and samples for response purposes to be more important, more time critical and more acceptable with fewer protections, particularly in emergencies, as illustrated by the following quote:

> *So if there's an outbreak like Ebola really, holding back identifying information like the breakout is in a certain village near a certain forest or near a certain water point or a certain migratory route and this can help the public health team that's reacting to the problem. I think honestly that's stupid. That's a different situation so* ***I try to differentiate care from research****. But if it's for research sake and just telling the stories, why not aggregate the data and tell the stories?*
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**

4.2.3.3. Qualitative data

Interviewees discussing qualitative data referred to a greater reluctance to share qualitative data – one interviewee discussed her own reluctance to share qualitative data she had collected, while another noted a general reluctance to share qualitative data. The reluctance was considered to stem from the possibly more personal nature of qualitative data, which may provide detailed

descriptions – although the idea that quantitative data can also be very personal was noted in section 4.1.2.3.

The following interviewee described undertaking life history interviews with people who have viral hepatitis, which is transmitted through varying routes including injecting drug use. In these interviews research participants were being asked to divulge very personal information. She emphasised the trust that is needed for her participants to share such details of their life with a particular researcher, and said she would feel that sharing this data would be violating this trust:

> *…And* ***we made the case that we did not want to do this for ethical reasons because they were life history interviews. It was actually impossible to anonymise them without rendering them completely meaningless.*** *So with qualitative interviews and with particular life history interviews, if you just take out the names and change people's names, change place names, that is a first level anonymisation that you might do when you're cleaning a transcript. There's so much, a wealth of data. People stories, their family makeup, they're talking about their schools, they're talking about their trajectories that you can't guarantee that somebody won't be able to piece together who that person is. I guess there's more than that. It felt like you'd entered into a relationship of trust with these people. They're from very a marginalised community. They're often engaged in criminal activities. For me, it's important to be able to win their trust, gain rapport, create a connection. And then you hear those stories. It's very much about the dynamic.* ***It felt like a bit of a violation of their trust to just go and dump their stories somewhere. Allow other people to see it.***
>
> **Interview 27, Sociologist Researcher, Oceanian, Research Institution, Europe**

A further notable aspect of the above quote is the suggestion that sharing is in some way dumping information. This suggests sharing is perceived as an undesirable or careless act, and contrasted with not sharing, which was presented as the ethical position. Although this particular interviewee is a qualitative researcher, and discussed sharing specifically around this kind of data, she did suggest that there may be similar concerns with quantitative data.

4.2.3.4. Aggregate data

As with sharing data and samples collected in different ways, a further arising sub-theme was the sharing of data and sample types that are held in different ways – specifically noted in this analysis

was the distinction between individual level and aggregate data, and anonymised and identifiable data and samples.

Interviewees were generally unconcerned about the sharing of data stored in aggregate form, in contrast to individual-level data where there were concerns around confidentiality, and potential harms for the individuals from whom the data or samples were obtained. This quote shows this contrast – it is the same interviewee who discussed qualitative data sharing above, now discussing data from research involving the analysis of urine samples:

> *That's not an issue for me and I have no issue with sharing raw data on what percentage of people and demographics had what their proteinuria, how many ever or referred to specialist. I have no problem sharing that sort of data.*
>
> **Interview 27, Sociologist researcher, Oceanian, Research Institution, Europe**

The aggregate nature of the data was seen to de-personalise them, and therefore the interviewee did not have the same concerns about sharing.

Although some did note the potential for populations to be stigmatised on the basis of work of shared data or samples, there was not discussion more specifically about the potential for group-level harms from aggregate data. Some interviewees did note some wider issues around data sharing more broadly that would also relate to aggregate data, such as the potential political sensitivity of the topic to which the data relate that may impact on a country's reputation, for example by revealing the circulation of an infectious disease.

4.2.3.5. Anonymised, de-identified and identifiable data and samples

The sharing of anonymised and identifiable data or samples were unsurprisingly viewed very differently from each other. Anonymisation was seen as a way of sharing data with fewer protections, as individuals cannot be identified. Interviewees on the whole were much more concerned about the sharing of identifiable data or samples as it was felt this may pose a risk to

those from whom the data or samples had been collected. As a result, interviewees considered the need for consent by those from whom the data or samples were obtained for sharing, and the protections required for the data or samples as much more important for identifiable data. What is notable, however, is that this for the most part was grounded in a concern around harms from sharing identifiable data, rather than a sense of wrongness of sharing the data without permission, irrespective of the potential for harm. This interviewee, however, presented a contrasting view that consent ought to be taken to share data, irrespective of anonymisation, as it was a part of respecting the research participant or patient's autonomy:

> *And it is not only forming risk [by anonymising data]. Yeah in ethical principles the corner…is to respect one's autonomy. To respect consent. That's why we are taking informed consent. We take informed consent because we respect the patient or participant. See, if we are sharing the data without letting [them] know about this process that is not respect for persons.*
>
> **Interview 9, Pharmacologist and Principal Investigator, African, Research Institution, Africa**

Some interviewees considered the importance of anonymisation, prior to sharing data, to vary by the type of disease the data relate to and the stigma associated with that disease, which will be discussed more in section 4.3 below.

One issue raised in this study was the level of anonymisation needed for wider sharing through a repository, and whether it needed to be different, with concern around the lack of clarity for this, as discussed by this interviewee, an epidemiologist based in Europe:

> *The only thing that the research team said was, "don't make the different parts of [city] identifiable." I had to figure out all the rest of it. The funder didn't know that was going on. It was all new. I was working out, step-by-step, what I thought was a sensible thing to do. There was no protocol. There wasn't a plan. I'm pretty confident that data set is safe to use, but it easily couldn't be and nobody would know. So I think there are data sets being made available that probably are not…have not been checked in that way. We know how to anonymise data. We've been doing it for years and years.* ***If I say a data set's anonymised and someone else says it's anonymised, I'm pretty comfortable that we mean the same thing by that. But if I say it's safe for putting in a repository and someone else says that, I don't know if they mean the same thing because I don't think there are rules.*** *Or, if there are rules, I don't know if they've been sufficiently thought through to think about all the stuff I was just talking about, this one in [city]. I was making it up as I went along.*

Interview 2, Epidemiologist Researcher, European, Research Institution, Europe

The quote illustrates the way that for many interviewees sharing is again perceived to impact the typology of data and samples (as with research and non-research data and samples), in the sense that 'anonymised' is considered to mean something potentially different when sharing through a repository. The interviewee also seemed to be expressing discomfort at the lack of safeguards to check if the anonymisation was appropriate.

A concern raised around sharing data labelled 'de-identified' was the potential to re-identify individuals:

> *I think at the moment that's a stage of, I'm at the stage where knowing that you don't know. I'm at the Socratic stage of knowing that I know nothing. Because when you look in the literature like I have a little bit recently. Then you learn that actually people have tried and sometimes succeeded to, not only to de-identify what supposedly was de-identified data, they've managed to crack the, and they've re-identified it.*
>
> **Interview 20, Medical Statistician and Epidemiologist, European, Research Institution, South America**

The above quotes show that while interviewees were often giving a recognisable account of de-identified or anonymised data being more straightforward to share than identifiable data, some also considered that the process of sharing may require more robust processes of anonymisation or removal of identifiers, and seemed uncertain about what these are.

4.3. The infectious disease context

The discussion in this chapter so far has focused on the types of data and samples being shared, and the way in which they may be shared. A further clear theme emerging in the analysis of the typology of data and samples relates more specifically to the infectious disease context, and how the type of infectious disease the data or samples relate to influenced perceptions on whether and how sharing ought to be done.

The different types of data and samples from the infectious disease contexts discussed were:

- Outbreak and non-outbreak
- Diseases with associated social stigma
- Diseases with associations with poverty

The sections below highlight some discussions around these data and sample types.

4.3.1. Outbreak and non-outbreak situations

Interviewees made a distinction between data and samples from outbreak and non-outbreak situations, due to the potentially rapid nature of outbreak response and the work that could be done on shared data and samples to facilitate the response. This led to varying perceptions about the implications for sharing them. Some interviewees, of varying backgrounds, suggested that data and samples related to outbreaks had more justification for sharing than those from non-outbreak situations. The interviewee below discussed her observations during the Ebola outbreak, and felt that the people were willing to go to great lengths and disregard rules in order to curb the outbreak and help with recovery:

> *From what I observed when we had the outbreak, even the patients, people were … Almost everybody, even the IRBs [Institutional Review Boards] were ready to break doors and windows and rules to move everything quickly, to bring recovery, bring information for prevention and just bring medicine. Just encourage care basically.*
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**

The 'outbreak' acted as a label on the data, with implications for the typology of data. The same interviewee considered that sharing identifiable data with the public health team was permissible in these circumstances:

> *So if there's an outbreak like Ebola really, holding back identifying information like the breakout is in a certain village near a certain forest or near a certain water point or a certain migratory route and this can help the public health team that's reacting to the problem. I think honestly that's stupid.*
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**

The quote shows that the public health response is considered here to be a priority, above maintaining confidentiality.

Data and samples related to emerging infections such as the Ebola virus were suggested by some interviewees to be of particular importance, because they related to a disease group considered a global health threat, and with relatively little knowledge. However, interviewees also discussed sharing data and samples related to emerging infections as challenging, due to the short timescales and potential chaos associated with outbreaks:

> ***You need to make a distinction between emerging infections and other infectious diseases****, because I think it's a very important issue. The emerging infections are for definition infections where there is a gigantic knowledge gap still remaining and potentially a lack of medical counter-measures. But it's not only the medical counter-measures that count. There's also other super important knowledge missing regarding transmission, regarding perception, regarding infection control measures, for example. I mean and I could go on, the list is quite long. And* ***therefore, if these emerging infections are coming, like Ebola, exclusively in the form of epidemics you only have a limited sort of data and time to make use of whatever you can grasp that could help you solve a number of these questions that you might have****.*
>
> **Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe**

As with interviewee 10, there is a sense of urgency clearly conveyed in this quote, and a need to use data as best as possible to manage the outbreak.

Some interviewees considered cautious sharing to be more appropriate in outbreak situations. This was due to concern about the chaos and fear surrounding outbreaks, and the potential to add further confusion if there is misinformation as a result of widespread sharing of data that are not quality assured:

> ***I don't think in an outbreak situation, I think for a start then everything's kind of a mess and everything's provisional and a lot of things are not confirmed. I don't think you want those raw data being somehow, you know, just open to everybody*** *yeah. Obviously in the Ebola, then from what I gathered, it was a complete scramble and even within the same institution people not knowing what other people were doing. But I'm not sure that that would come under, depends how you consider data sharing yeah. I don't think that requires what you might call it data sharing but it's more the classical kind of collaboration, I think it's more people collaborating. And agreeing to, hopefully, toward common systems, but not in the sense of having in some, going in real time to some open repository.*
>
> **Interview 20, Medical Statistician and Epidemiologist, European, Research Institution, South America**

The quote relates to the distinction identified above in section 4.2, between sharing with collaborators compared to a repository, and the greater level of concern from some with respect to sharing through a repository with people who may be unknown to them.

A small number of interviewees questioned the distinction between outbreak and non-outbreak data, noting that either way there were people being affected by the infectious disease in question, whether or not it was part of an outbreak – suggesting that timely data sharing is important and justified either way. This interviewee, a laboratory scientist with experience across a range of infectious diseases, explained her view that while it may be better where there is time to ensure data are quality assured, delays in sharing may cause harm whether or not the infectious disease in question is part of an emergency situation:

> *…**I imagine if you think data sharing is a good thing, and enables research to happen more quickly, then in theory, it shouldn't matter if it's an emergency or not.*** *But I guess it depends, if there is an issue over quality assurance, having the time to peer review thoroughly, if it's not an emergency, then you can see why that might be preferable. Because if you don't think that anymore harm would come from the delay, then, I don't know. But **you could probably say that all delay could cause harm, because ultimately, you're trying to prevent infection. Whether it's an emergency or not, you still may cause harm by not publishing, not sharing data.***
>
> **Interview 1, Laboratory Scientist, European, Government Agency, Europe**

4.3.2. Social stigma of disease

Data and samples relating to a socially stigmatising disease were discussed by a number of interviewees as a distinct type. One reason some diseases were considered potentially more stigmatising than others was to do with their mode of transmission, for example sexually transmitted diseases, or diseases transmitted through a behaviour that is criminalised such as injecting drug use, with the levels of stigma differing by context. Where diseases were noted to have a higher level of social stigma, the associated data and samples were considered by interviewees to be more sensitive and requiring a high level of protection, thereby constraining sharing. This is illustrated by the quote below which goes on to suggest there are risks to an individual if data are shared for a stigmatising disease like HIV. The interviewee went on to explain the care that needs to be taken with the way data are kept confidential, to protect the identity of individuals.

> *Different infections have different sort of status with regard to confidentiality and have different meaning to people in regard to confidentiality. Something like HIV, there remains major stigma associated with HIV, number one. But number two, because we work in molecular epidemiology ... We work in areas, we could start to suggest about transmission chains, then there also becomes an issue of disclosure of who may have infected who.*
>
> **Interview 15, Doctor and Scientific Director, European, Research Institution, Africa**

In contrast to the above, diseases seen as commonplace, affecting the bulk of the population in certain regions of the world, such as malaria, were noted to be less stigmatising, and the associated data and sample were seen as less sensitive:

> *Malaria is really quite easy compared to almost all the rest of these, for a long time anyway it's been happening everywhere. Not everywhere but in huge parts of the Earth, everybody gets it, nobody feels stigmatised or anything for having malaria and so the emotional content is lower. Not that nobody dies of malaria but it just doesn't have the same flavour. For diseases that are always with us as it were.*
>
> **Interview 16, Scientific Director, North American, Research Institution, North America**

Of course, it is possible that a stigmatising disease may come in the form of an outbreak, leading to an impetus to share, but also a reason for caution. This was discussed in the context of Ebola in this quote:

> *But the other thing is, which is always a worry for me, a recent paper actually publicised…* ***evidence that there are super-spreaders of Ebola. Some people who contribute far more to ongoing spread than others for whatever characteristics that is. So there are, again, there is sort of a little bit of a blame sort of culture that can develop in terms of some people's perceptions.*** *If you find that your brother who has died actually contributed excessively to other people's infections then, of course, that has a number of implications. I think Ebola, there remains a cultural ... There are sensitivities there to identifying who may have infected who in terms of transmission.*
>
> **Interview 15, Doctor and Scientific Director, European, Research Institution, Africa**

Taking this quote in the context of the discussion above in section 4.3.1 on outbreaks, it shows the different considerations that promote and constrain sharing. The Ebola outbreak was discussed as one in which it may be more ethically permissible to share identifiable information for the public health response. The quote above sheds a different light on data related to the same outbreak, discussing a concern that people could be identified as 'super-spreaders', that is, people who have infected a number of others.

4.3.3. Diseases with associations with poverty

The nature of the population that is usually affected by a particular disease emerged as a further factor informing the way interviewees discussed data and samples from the infectious disease context. In particular, neglected tropical diseases (NTD) such as VL were noted to affect poorer people. Those discussing NTDs suggested a greater potential for benefits from sharing, given the low levels of evidence available in relation to them:

> *Especially in neglected tropical disease areas where limited clinical trials have been conducted, and the data that is available from such trials are quite less. So if you are able to consolidate and bring together data from a variety of studies globally, then at least you'll be able to have some data that can make sense, statistically. Unlike diseases like malaria, where lots and lots of clinical trials have been conducted,* ***neglected tropical diseases there are very few trials that have been done, so if you***

can share data that is available, then it would make it easier to be able to look deeper into available data, and probably to accelerate research.

Interview 19, Data Manager, African, Research Institution, Africa

The same interviewee also highlighted the lower political weight given to data relating to diseases that predominately affect people in poverty, going on to suggest that governments would place much lower emphasis on the importance of sharing such data compared to outbreaks affecting a wider segment of the population:

…an outbreak of VL might not give any government a headache, one because ***VL affects poor people in very rural areas in most countries, and this in many cases that governments are not as concerned about all of this.*** *But whenever there are maybe some other outbreaks that affect majority of the population, then maybe, regardless of the social class of these people, then you know, this will be a big concern to the government. In such cases, I mean, they will not want to think about some regulations, you know, stopping health officials, or even the scientists for gathering the best information they need to can then contain such kind of outbreaks. In those cases it will be natural, yes, to collect data as much as possible.*

Interview 19, Data Manager, African, Research institution, Africa

This leads to concerns with justice, if data and sample sharing is de-prioritised for some groups, and the potential benefits are not realised.

Others considered the distinctiveness of data or samples related to diseases that predominantly affect those living in poverty, and the sharing of them, to stem from the difficult social circumstances of those from whom the data or samples were collected. This related to a concern that the people from whom data or samples were collected were particularly vulnerable and in need of additional protections, or that they may not be aware of the implications of data or sample sharing and so truly informed consent would be difficult to obtain. The fact that the data or samples related to infectious diseases became secondary to some interviewees, as highlighted here:

I think it's not really the fact that it's infectious diseases. It's more settings where we are working, in sub-Saharan Africa, in Asia. Countries where we don't have so many resources to run serial research in …. I think it's more that than the nature of the disease itself.

Interview 7, Clinical Trials Manager, European, Research Institution, Europe

4.4. Conclusion

This chapter has examined the ways in which key morally and practically significant differences between types of data and samples, and between different modes of sharing emerged in my analysis. The analysis shows that 'data and sample sharing' is not one thing, but includes a number of different types of data and samples and a number of different ways of sharing them. There were two clear distinctions that emerged with respect to mode of sharing. Firstly, whether sharing was happening between known or unknown groups of people. Secondly, with respect to data sharing with unknown people, whether access to the data was open or managed – that is, whether there were assessed criteria for accessing them (sample sharing was never discussed as open). The term sharing was also itself contested by some interviewees, who considered that the reciprocity element was lacking when sharing to a repository.

It is clear from this chapter that interviewees generally considered different types of data and samples to be more or less morally permissible to share, primarily grounded in concerns about what the data or samples could reveal about the people from whom they had been collected, and the potential for harm as a result of that. Although an important observation here is that different interviewees also had different perceptions around the degree of distinction and impact on the permissibility of sharing. Data sharing and sample sharing were also generally considered distinct from each other, with some different ethical considerations related to sharing them.

Finally, the infectious disease context was itself considered to be diverse: sometimes, but not always, an emergency context, and including a number of different disease types. Some of these diseases were also noted to predominantly affect certain populations, such as those who were poorer, and some diseases were noted to be more stigmatising than others. All of these distinctions

with respect to the type of infectious disease again altered perceptions around what was permissible to share – although again it is notable that there was not consensus among interviewees around this. For example, identifiable data were generally seen as less permissible to share, but for some interviewees this was changed in the context of an outbreak, where some felt that it was then appropriate to share them for the public health response. Others, however, felt that outbreaks ought to lead to more guarded sharing due to the chaos and fear.

The next chapter will pick up on the theme of sharing between diverse groups of people – the seed of which has been planted in this chapter.

5. Organisations and professions, and their differing priorities

The previous chapter demonstrated that for those working in the field, there are a range of different categories of data and samples, each of which is viewed as being associated with different ethical implications for sharing. The focus of this chapter is on the types of organisations and professions involved in infectious disease data and sample sharing, and their varied priorities. This was another major theme emerging from the analysis.

The first section (5.1) outlines the roles of the different organisations and professions discussed in the interviews, in relation the data and sample sharing. The second section (5.2) focuses on the perceived priorities of these different organisations and professions. The final section (5.3) more specifically focuses on the identified priorities in relation to low- and middle-income country (LMIC) contexts.

5.1. Role of organisations and professions

Interviewees identified and referred to a large number of organisations and professions involved in infectious diseases data and sample sharing (shown in Figure 5.1, below). Some were characterised as data or sample contributors, while others are seen as data or sample users. Some may have both these roles. Other organisations were seen as relevant to sample and data sharing practice as funders or policy makers. Interviewees discussed their perspectives on the roles that certain types of organisations and/or professions do, or ought to, play with respect to data and sample sharing, and critiques of how well the role was fulfilled. In what follows, I discuss the major organisations referred to by interviewees. I begin with the World Health Organization (WHO).

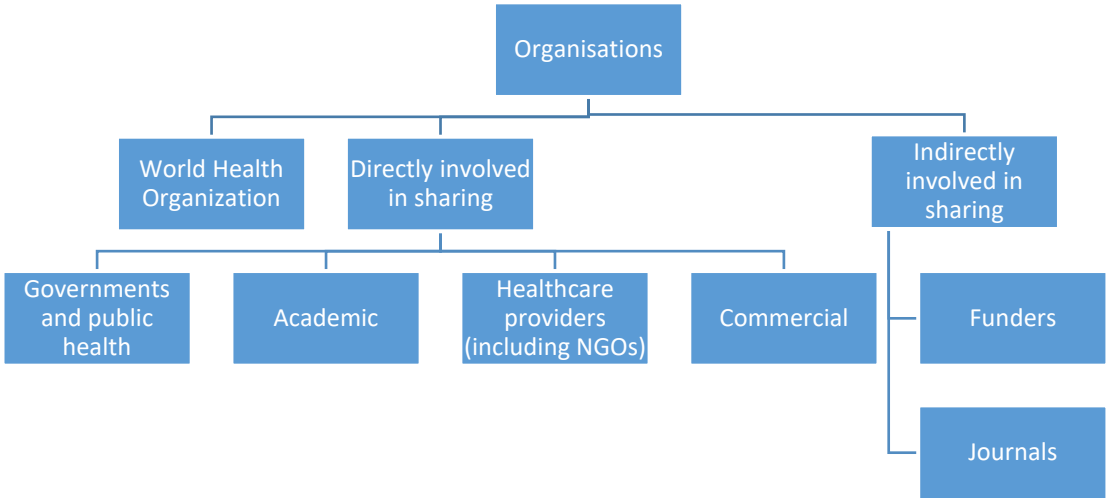


Figure 5.1: Diagram of organisation types involved in data and sample sharing from the interview analysis

5.1.1. The World Health Organization

5.1.1.1. The WHO as a leader and source of guidance

The WHO, as the agency of the United Nations, was considered a leader in the global response to infectious diseases by interview participants. The International Health Regulations (IHR) 2005, discussed in Chapter 2, place an obligation "…on States Parties to notify WHO of events that may constitute a public health emergency of international concern according to defined criteria" (94), thereby placing the WHO as a central actor within international data sharing activities. The WHO also produces guidance on activity related to infectious diseases data and sample sharing, and is the convener of the Research and Development Blueprint, which includes work to develop norms and standards with respect to data and sample sharing during public health emergencies (139). As one interviewee (interview 16) stated, when discussing a sample sharing initiative "…WHO was very involved and that was important as well because their influence and whatever is key to much of what goes on."

While this was considered the appropriate role for the WHO, some interviewees felt that the WHO had fallen short of fulfilling it. One interviewee involved in developing data sharing platforms discussed this:

> ***It would be ideal to have very strong authority that was providing leadership on this and WHO would be the automatic, sort of, reflexive body that would do that. And that hasn't happened to date****, I think they're trying but because of the political process that is the WHO, it's incredibly slow to develop. Not to say they haven't tried to be supportive, but I think that there's a huge gap there that would resolve a lot of these issues if more time, effort and attention were paid to this topic at the WHO level…*
>
> **Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe**

The interviewee clearly perceives that the WHO had not provided the sort of leadership that she considered necessary, suggesting that the WHO was slow to act and not sufficiently engaged with those who were directly working on data and sample sharing.

5.1.1.2. The WHO as a locus of knowledge sharing

Alongside the WHO's function as a leader and source of guidance for the global health response to infectious diseases, they were noted to have a more direct role in data and sample sharing, acting as a locus for gathering and disseminating information. This is also a part of IHR (2005) requirements, as discussed in Chapter 2. This role is described by one interviewee who had experience working with the WHO during the Sudden Acute Respiratory Syndrome (SARS) outbreak:

> *…around SARS in particular we [WHO] did ... because, about SARS, the gathering of information was absolutely critical. From the WHO we were convening all sorts of networks, just the people who were seeing patients and were doing things, on the phone with them all the time. Saying, "What are you seeing? What are you finding?" Sharing things amongst the epi people, the lab people, the marketing groups, etc. That was working very well but there were all sorts of other people involved who were out there, who were gathering information.*
>
> **Interview 18, Doctor and Policy Maker, European, Government Funding Agency, Europe**

In contrast to the quote in section 5.1.1.1, this quote shows a perception of the WHO as a successful locus of data sharing during an international outbreak.

5.1.2. Organisations directly involved in sharing

In addition to the WHO, the following organisations and associated professions were discussed by interviewees as those that are often more directly involved in sharing – for example, as data or sample contributors, platform conveners, or users of shared data or samples. Some of these organisations will also have other roles, for example policy making or advocacy.

5.1.2.1. Governments and public health professionals

Interviewees generally discussed the role of governments in relation to their public health function, as presented below.

Using knowledge for infectious diseases response

Public health organisations were primarily discussed as responders, using knowledge gained through use of shared data and samples to respond to infectious diseases. There were a wide range of potential uses of knowledge derived from shared data, such as resource allocation, logistics, and contact tracing, as discussed in this quote:

> *I suppose* ***anything you learn about during an outbreak, epidemic, whatever term you want to use, can affect or not what is being done in the response.*** *Either as public health measure, where you put the resources, where you send the teams, who does what, how many people have to be working on what, who to follow up in the community, who might not have been exposed, how to understand what the patterns of transmission are, so if necessary change your interventions. If there's something that relates to clinical observations or clinical care, obviously this has to be understood and transmitted to the people who are in charge of clinical care, infection control.*
>
> **Interview 18, Doctor and Policy Maker, European, Government Funding Agency, Europe**

This role was seen as complementary to the researchers' role in producing scientific outputs that inform the public health response, using shared data and samples.

Public health organisations as data compilers

Public health organisations - which are often government organisations - have a role in synthesising data and analysing samples for surveillance purposes, and sharing this with organisations such as the WHO. Surveillance is described by the WHO as "[a]n ongoing, systematic collection, analysis and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice" (140). Surveillance data can be compiled from notifications of cases of an infectious disease, and there may be surveillance tools used to gather additional information in relation to certain infectious diseases. Interviewees in this study did not discuss the role of public health organisations in collating and analysing data. Public health was usually associated with the response to infectious diseases as outlined above.

5.1.2.2. Academic organisations and researchers

In the previous section we heard about the role of public health. This section explores the diverse roles of academic institutions and researchers in relation to data and sample sharing.

Producing more science

Researchers were primarily considered to have the role of producing more science to better understand infectious diseases, and to develop innovations to respond to them. Research was viewed by a number of interviewees, often researchers themselves, as being an essential part of the infectious disease response – and data and sample sharing, through facilitating more research, was therefore also deemed essential.

Academic researchers were discussed as both data and sample contributors, and users of shared data and samples. Those involved in conducting clinical trials considered data sharing to be a means of increasing sample sizes, giving more statistical power and enabling certain sub-group analyses that might not be possible in an individual trial. This would then give more evidence around areas such as the treatment of the disease in question:

> *The main reason is to be able to increase the number of patients, and somehow be able to get some more information. I mean, it's always a limiting factor, the number of patients in a clinical trial. You would always want to have many patients but it's very costly and so on, so it's always targeted to one objective. And you have usually plenty of other secondary endpoint questions where you don't have the power to really get the final conclusion. And* ***being able to share with others could probably help identify something that would be minor for a clinical trial but could be very important for treating the patients and for the outcome, the knowledge about the illness.***
>
> **Interview 7, Clinical Trials Manager, European, Research Institution, Europe**

Facilitating data and sample sharing

A further role of academic institutions described by interviewees in this study was facilitating data or sample sharing through developing data repositories and biobanks. The following interviewee, discussing her experience in setting up a data sharing platform, explained that academic institutions provided an appropriate base, as those engaged in data sharing were predominantly academics, and there were practical benefits to housing data sharing within academic institutions:

> *We decided that to be an NGO was too complicated, you need to do your own legal and HR and financial and we don't know how to do this so that seemed like not a good idea. We seriously considered being somehow attached, embedded, or whatever within WHO but decided that in that case it would not work because we needed the flexibility, especially in terms of hiring. The rules at WHO are, I don't know, beyond me I'm sure. So that one after much debate was decided to not be chosen. Then we said, "Okay, academic. Okay, where?" And the decision to choose [university] was really a practical one…*
>
> **Interview 16, Scientific Director, North American, Research Institution, North America**

The quote also shows that the WHO is perceived as inflexible by this interviewee, a potential further barrier to fulfilling its role in data sharing during a crisis.

5.1.2.3. Health care providers, including non-governmental organisations

Health care organisations were seen as involved in data or sample sharing primarily as contributors of data and samples collected in the course of clinical care. Within the context of infectious diseases emergencies, particularly in low income countries, a significant amount of health care will often be provided by non-governmental organisations (NGOs). Low income countries experiencing an

outbreak may have fragile health systems, with issues with respect to system governance, funding, workforce, infrastructure and supply chains leading to a system that could easily be overwhelmed with increasing demand for healthcare during an outbreak. This was clearly an issue during the West African Ebola outbreak in 2014-16[11], when a number of NGOs and overseas governments provided Ebola Treatment Centres (141). In the course of healthcare provision and undertaking or collaborating in research at the treatment site, large volumes of data and samples would be gathered.

NGOs also have other roles in relation to infectious diseases, including as funders, and through public health, research, community engagement and advocacy activities. Within this study, with the exception of NGO funders, NGOs were discussed as data contributors, sharing data collected as a part of delivering health care and conducting research during an infectious disease response.

Some NGOs were suggested by one interviewee (interview 26), who was based in an NGO, to potentially not have much knowledge about data sharing, and research. In particular she discussed smaller NGOs, stating "…if you're a small NGO and you're being asked to share, do you even have the people who even understand anything about that?"

The work of data sharing may, in some cases, not be related to their core area of expertise. The interviewee below, who had experience working with an NGO during an outbreak, discussed the reluctance of the NGO to share data that had been gathered through providing healthcare during an outbreak. She suggested that the lack of sharing was due to a lack of familiarity with data sharing:

> *We didn't put the [anonymised] dataset in the repository because [international NGO] didn't want us to. They were very, very uncomfortable with us sharing data. If it had gone in a repository, I would be okay with that…I'm not sure I ever understood really why it was.* ***I suppose it was an organisation***

[11] There has been a great deal written about epidemic preparedness and fragile health systems. Two concise perspective on the fragility of the health system in the three most affected countries during the West African Ebola epidemic in 2014-16 can be found here: Boozary AS, Farmer PE, Jha AK. The Ebola Outbreak, Fragile Health Systems, and Quality as a Cure. JAMA. 2014;312(18):1859–1860. doi:10.1001/jama.2014.14387; and Kieny MP, Evans DB, Schmets G, Kadandale S. Health-system resilience: reflections on the Ebola crisis in western Africa. *Bull World Health Organ*. 2014;92(12):850. doi:10.2471/BLT.14.149278

that was quite new to doing health work at all and wasn't used to doing medical research. *I mean, some NGOs are more familiar with it. It was all brand new to them, so I think that made them uncomfortable with [the] whole thing from the start.*

Interview 2, Epidemiologist Researcher, European, Research Institution, Europe

The quote also shows a dissonance between what the individual is happy to do – in this case to share the data - and what the organisation mandates, which is in this case not to share the data.

Other healthcare providers discussed were government providers, considered to be a part of the government public health role noted above. Private for-profit healthcare provision was not discussed.

5.1.2.4. Commercial organisations

A number of commercial organisations may be involved in data and sample sharing during infectious disease research and responses, such as technology companies, pharmaceutical companies, and transport companies. Pharmaceutical companies were the main commercial organisation discussed in this study, by a small number of interviewees. They were noted to be either data contributors, or users of shared data and samples. Pharmaceutical companies may be involved with drug and vaccine research and development activities.

Pharmaceutical companies as data contributors

The data collected from pharmaceutical research including clinical trials was noted to be of potential use more widely, to pool together with other data. The following interviewee discussed his experiences when developing a data sharing platform. He described pharmaceutical companies as initially a reluctant participant in the data sharing, as they were concerned about the potential negative impact on their product. He explained that over time they became a useful contributors of data, and particularly helpful with understanding the data they had shared:

> *It has become fairly straightforward and standard as a relationship with pharma because what we put in place is a clear process. There's rules, terms, process and pharma they know what's going to happen, they know what's going to be the rules and they know what's going to be the outcome. They know where they can intervene and when they can't. And they are very comfortable with that. And actually they became probably the groups who were most interested in the analysis of the data and gave us, actually, really interesting clues about the data because of the internal understanding of their data.*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

Pharmaceutical companies as data users

Pharmaceutical companies were also seen as potential data and sample users, using them to help further research and development activities which may be needed for the public health response, as discussed by the interviewee below in relation to developing an Ebola data sharing platform:

> *If it's a drug company who intends on developing therapies for Ebola, I think that's absolutely legitimate, absolutely the purpose of the platform and we would just have to ensure, as we would with a university that was developing research, that the outputs of that research were accessible to the populations from where that data came.*
>
> **Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe**

An issue she raised, which is explored further in subsequent sections below, is the distribution of benefits that arise from such research and development.

5.1.3. Organisations indirectly involved in sharing

Having focused thus far on organisations directly involved in infectious diseases research and responses, this section focuses on funders and academic journals – two organisations (or functions of organisations) often discussed by interviewees in relation to data and sample sharing. These organisations were not discussed as direct contributors or users of data or samples, but as a part of the wider data and sample sharing organisational landscape.

5.1.3.1. Funders

Funders as promoters of data sharing

Funders were discussed as promoters of data sharing (with less discussion in relation to samples). They were also seen integral to the academic system, and therefore having influence and even mandating that data sharing is done.

One interviewee, who has a background working for funders, suggested that funders and journals were key to data sharing taking place. She discussed the Global Research Collaboration for Infectious Disease Preparedness (GLOPID-R)[12], which is an international network of research funders (142), as an important driver for data sharing during public health emergencies. She noted that funders and journals could also be barriers to data sharing moving forward if they do not support it, highlighting the influence that they are perceived to have on data sharing:

> *A lot of work has been, let's say, given a further motivation by the GLOPID-R ... development of GLOPID-R, their establishment of a working group, because* ***it's my view that the funders hold the key to making this [data sharing] happen or not.*** *I think there are two big aspects that were either obstacles or just not making it happen and one of them was a whole variety of issues around whether the funders support this, and the other issue is of course about the commercial scientific publishers.*
>
> **Interview 18, Doctor and Policy Maker, European, Government Funding Agency, Europe**

This role in promoting data sharing wasn't necessarily viewed positively by interviewees. The following interviewee, who is involved in developing data sharing platforms, considered the funders as being heavy-handed in their approach. He suggested it came across as a top-down initiative that wasn't sufficiently engaging with the people from the ground:

[12] GLOPID-R was developed in 2013 and describes itself as "bring[ing] together research funding organizations on a global scale to facilitate an effective and rapid research of a significant outbreak of a new or re-emerging infectious disease with epidemic and pandemic potential". They have a data sharing working group, who have released some publications on data sharing (including a roadmap, and principles) subsequent to the collection of the interview data for this study. See: https://www.glopid-r.org/ (last accessed 1 October 2020)

> ***I think there are some work ongoing, mostly by funders. But, this is ... It's not very inclusive of the community and I don't think that they are…I think that they are over-simplifying the problem today.*** *I don't see them being realistic about the barriers that would have to be overcome. They think that simply by putting a rule as a funder that will solve everything. I think that they are not very sensitive of the perception of risk, from the countries' perception of risk for the individuals, academic battles that will occur during these events…*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

This quote suggests that it is not necessarily a case of groups who object to the approach taken by funders as being 'against' sharing, but that those who are themselves involved in promoting sharing have differences in what they consider the appropriate approach to doing so

Another interviewee also expressed concerns with the way that data sharing was being promoted. She discussed the role of funders convening meetings about promoting data sharing, suggesting that as the funders were generally based in high-income countries (HICs), those from HICs were disproportionately represented in discussions, and devising rules:

> *What is now happening a lot, and it's a bit understandable, but it's still disturbing, there is a lot of meetings with funding agencies. First the funding agencies tend to be from high-income countries, no surprise. They want to be prepared, they're pushing for us to do data sharing, and all sorts of stuff, and it's fine. But when they discuss these issues, and what could be the rules.* ***What is disturbing is that it most certainly occurs in a majority countries from low-income countries, or maybe mid-income countries, and the guide for discussing are all from high-income countries.*** *That's a bit disturbing. I do understand that they're going to be the ones funding it most probably, because the low-income countries won't have the money to fund ...*
>
> **Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe**

The concern looks to stem from the mismatch between the burden of infectious diseases being disproportionately in LMICs, and the guidance from funders being produced in HICs.

<u>Funding for data sharing</u>

Interviewees who developed platforms, and from funding organisations, also discussed the role of funders to fund data sharing. This interviewee, a doctor based in a European funding organisation,

stated that where funders were requiring data to be shared, they should also provide funding for the data sharing itself:

> *One of the questions that has been raised as a funder, if you are enforcing, if you are saying to the organisations you're funding that there should be pre-publication, sharing of data, particularly with relation to epidemics, then you need to be funding organisations to enable them to share their data and have mechanisms to share their data the appropriate way as well.*
>
> **Interview 14, Doctor and Policy Maker, European, Government Funding Agency, Europe**

Some researchers felt that despite the rhetoric about its importance data sharing was inadequately funded, and as a result they were expected to prepare data for sharing without extra funding. These insights are at odds with the discussion above that portrays funders as key advocates for data sharing:

> *You know, I think people haven't stepped up to the plate with respect to funding this [data sharing], maybe because they don't understand it. Maybe because they haven't seen the value of it. Maybe because it takes a really long time to change things in organisations like these, which are massive structures.*
>
> **Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe**

5.1.3.2. Journals

Journals as promotors of sharing

Journals were also discussed as promotors of data sharing, joining funders to release statements to promote data sharing in public health emergencies, as noted in Chapter 2. The following interviewee discussed a requirement from some journals to share the data associated with the published research article:

> *We also have journals, publication journals that insist that before you ... it's a major article for publication then you have to make data that you're publishing about available in a data sharing platform.* ***Those are basically forces from external parties require that data is shared****.*
>
> **Interview 19, Data Manager, African, Research Institution, Africa**

Similarly to funders, there was also concern that journals were engaging in a top-down approach by requiring data to be shared. Some, such as the interviewee below, also felt that it wasn't necessarily made clear to researchers submitting papers how to fulfil the data sharing requirement:

> *Well, I think everybody has his or her own idea, [journal], for example there are many other channels that have followed suit, they say something like, "The data set used to actually produce this paper must be open."* ***The biggest problem I think with these, sort of, dicta from on high*** *is that they say that but they don't tell you what they really mean and people are not at all clear about how to do it.*
>
> **Interview 16, Scientific Director, North American, Research Institution, North America**

Taken together, these two quotes above from interviewees with different backgrounds show that data sharing can be perceived as a top-down imposition. They are also both involved in developing data sharing platforms and are therefore not 'against' data sharing.

The following interviewee also considered the model of journals in general – a model that is based on publishers charging for access to the articles[13] – not to be well aligned with the aim of data sharing. She considered that they were changing practices and publishing material that had been shared previously on pre-publication sites, but felt that the drivers of data sharing remained funders:

> ***Of course, their whole model is not founded on open access to information anyway, there is some limits.*** *That's changing a bit too. I think what is interesting is that now with the real growth in…pre-publication sharing sites, which are not data per say but they're publications address publications, which have not yet been submitted or have been submitted but not been proved for publication. And you still see some of this stuff make its way into good journals, so obviously I think the journals are coming to grips with this.* ***Possibly not at a stance until now but to me the real answer is not the journals. The real answer lies in the foundations or the science funders.***
>
> **Interview 18, Doctor and Policy Maker, European, Government Funding Agency, Europe**

[13] There are journals that are open access, or certain articles may be open access – the financing often then comes from those submitting the work paying a fee for publication.

This again suggests a perceived dissonance between the rhetoric around promoting data sharing, and the incentives in place that may work against sharing – although there is also a suggestion in this quote that this is changing.

Journals as quality assurers

Journals were noted by interviewees with a research background to also have a role in quality assurance and transparency. The process of peer-review was suggested to be one way that journals could scrutinise research outputs resulting from data sharing, thereby mitigating some of the concern that shared data or samples may be poorly analysed by others. This is the same as the role peer-review has for publications relating to the primary analysis of data or samples, and so the issue of data or sample sharing here is incidental, as it is about the wider system relating to the quality of publications from research outputs. This interviewee discussed the role of peer-review in quality assurance:

> ***I don't want data analysed by people who are incompetent, but you would hope that the, you know, peer-review would help with that.*** *Analysis is one thing, but making it publicly available is another. But then, can't really rely too much on that. Going to say, "we're going to put all the weight onto the journal to sift out what's good and what's bad." We know that that's tricky.*
>
> **Interview 2, Epidemiologist Researcher, European, Research institution, Europe**

The interviewee expressed concern that putting responsibility squarely with the journals to identify poorly conducted analyses may not be a reliable way forward.

The other quality assurance role discussed was that through mandating data related to a published study to be shared, it would enable others to check the analyses of that published study. The interviewee below discussed this, suggesting that this would make the process more transparent, as people can look at the underlying data:

> *There's issues with transparency. So journals and they claim to be able to access the data for a sense of transparency so people can look at it, can see on what these interpretations are built, and stuff like that. I guess there is … that it would enable more comparative work to be done.*

Interview 27, Sociologist Researcher, Oceanian, Research Institution, Europe

5.2. Organisational and professional priorities, and emerging tensions

One of the key themes emerging in the analysis presented so far in this chapter is that a range of organisations and professions are involved in data and sample sharing in the infectious disease context, and they are perceived to have different roles. This section will focus on the different perceived priorities that arise from the different organisations and professions involved. This overarching point is explained by the below interviewee, who discussed her own experience working in the response to the Ebola outbreak, highlighting the different organisations and professions coming together, and their varied cultures and values. These different organisations will be collecting data, and sometimes samples:

> ***So many times in working in [West Africa], it became very apparent that there were very different cultures. There's the NGO culture and the military and the [anonymised] guys and the researchers, and the NHS, and yeah ... Huge cultural differences, even between all the people who spoke English, even before you started looking at the [West Africans].*** *And yeah. I never realised how striking that was. I think because they were all, like yeah... The NHS, the NGO world, the military are large, they're driven by values but they're different sets of values. They're all things that people feel proud to be a part of and feel a sense of identity with. None of those are just jobs.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

The interviewee draws attention to her experience that such a plethora of organisations resulted in a gap in terms of their culture and understanding of each other's motives, creating difficulty with sharing data.

Some of the priorities explored in this section were considered to be particular to a specific organisation or profession, while others were seen as more cross-cutting priorities for a number of organisations. A notable feature of this discussion was the caricatures that sometimes appeared about different professions and what their perceived priorities are, which at times diverged from

the account that people from those professions gave of their own priorities. The main priorities emerging from the analysis were public health, academic, health care, reputational and economic, and commercial priorities. These are presented in turn in the sub-sections below.

5.2.1. Public health priorities

The public health priority within the infectious disease context focused on the ability of public health action to potentially save lives and reduce the spread of disease, with data and sample sharing suggested as contributing to this.

5.2.1.1. Saving lives and improving health

One prominent reason put forward by a number of interviewees from a range of backgrounds to share data and samples was to use data for the public good - to save lives and improve patient care in the context of outbreaks. The interviewees suggested that knowledge gained from the data being available to public health organisations for planning purposes, or through additional research being conducted on the data and samples, would improve the response to infectious diseases – either an ongoing response or a future response:

> …*it's absolutely clear that* ***no one institution can control and respond to an epidemic on its own*** *and through experience we've realised that actually* ***if we had information data a lot earlier it would inform our patient care and our policies a lot quicker and then we could bring an outbreak to an end quicker and hopefully ultimately save lives. And it is as profound as that****. I really do think it's as profound as that.*
>
> **Interview 22, Doctor and Group Head, European, Research Institution, Europe**

The quote also links sharing to speed, that it provides the potential to respond to infections more quickly.

As discussed in Chapter 4, data from outbreaks were sometimes viewed differently, and some interviewees gave greater or even an overriding priority to sharing data for public health action to save lives in outbreaks compared to non-outbreak situations.

5.2.1.2. Sharing early and rapidly

A notable theme from the analysis on sharing data and samples to meet public health priorities was speed – with interviewees frequently suggesting a need to share data early and rapidly, as stated in the quote above, and by the following interviewee:

> *...if these emerging infections are coming, like Ebola, exclusively in the form of epidemics you only have a limited sort of data and time to make use of whatever you can grasp that could help you solve a number of these questions that you might have. And* ***so you should take advantage of whatever material you have to increase your knowledge for the benefit of the future****, preferably deterrent is always better, but that in this case is not possible anymore. At least for the future patients and communities that will be affected.*
>
> **Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe**

This was put forward for the benefit of other people – people who may in the future be affected by the disease, alongside the current emergency being dealt with. Rapid sharing was something that was considered a particular requirement of the emergency context, to meet the public health demands of these situations – and, as noted by the interviewee below, potentially of less importance for non-emergency situations. In these cases, he suggests it would be better to take time to clean the data first:

> *I think for a lot of situations … I think there are also some situations where you could ask the question: Will anyone gain if this data is shared before it's published? No one will actually gain from a public health perspective, and it's actually more useful to tidy up the data, to clean it to avoid any confusion. So* ***I think there is something very specific about epidemics and the public health impact of sharing the data early.***
>
> **Interview 14, Doctor and Policy Maker, European, Government Funding Agency, Europe**

However, as noted in the analysis in Chapter 4 a small number of interviewees did question whether there was a special requirement to share early and rapidly in emergencies only, suggesting

that rapid sharing could also save lives in non-emergency situations as infections are still being transmitted.

5.2.2. Academic priorities

Discussion of academic priorities with respect to data and sample sharing was diverse. Some interviewees, many themselves researchers, suggested that an important priority for researchers was to contribute science that would improve the public health response to infectious diseases, thereby aligning with public health priorities. Others considered it a priority for researchers to protect study participants, which is discussed further in Chapter 6. A further priority, considered specific to researchers, was the pursuit of academic recognition.

Of note, while many people conducting research in the infectious disease context are based in universities, others are based in public health organisations, NGOs, or commercial organisations, and so the researcher profession is spread across different types of organisations.

5.2.2.1. Academic recognition

An often discussed priority was the incentives created by the receipt and distribution of academic rewards in relation to data and sample sharing by researchers who were participating in sharing initiatives. The main form of recognition discussed was being a part of publications, either though authorship or acknowledgements, and interviewees from a diverse range of backgrounds noted the importance of recognition for the academic career trajectory, with competition for promotion suggested to be more pressured in certain LMICs, as will be discussed further in section 5.3. This type of priority was said to both motivate sharing, and lead to concerns about sharing. Some interviewees suggested that researchers may be keen to participate in data and sample sharing so that they could gain publications (with some initiatives explicitly making this a part of the sharing process), whilst others suggested that researchers may refrain from sharing because of concerns that they will lose recognition related to data and samples they have participated in collecting when

others go on to publish further analyses. This could then impede the speed at which data or samples are shared, thereby conflicting with public health priorities.

Some interviewees outlined a variety of motivations for those who may be involved in sharing, generally in order to harness skills to generate more research and evidence, but also a potential motivation to become involved to gain publications:

> *I guess the main reason is, for most people, what I've seen at least that it's possible to do with certain analysis that is not possible within the consortium or within a certain person's institute and that could possibly give more evidence or really directly, also maybe improve patient care and the best case even for that patient from which a sample is taken in the most direct way. So I guess those are the main driving factors but yes, there's other reasons as well of course, just increasing publications and more academic rationale, not so much related to improvement of patient care but more related to improvement of scientific records.*
>
> **Interview 4, Pharmacologist Researcher, European, Research Institution, Europe**

This quote also shows the difficulties of generalising – that people may be involved with a range of priorities in mind, and the priorities can differ between individuals.

5.2.2.2. Maximising use of funding

A further priority, discussed predominantly by researchers, was to make better use of the funds that had been put into collecting the data or samples. Data and sample sharing was seen as a way to reduce the resources needed to get certain results and research outputs. This was highlighted by the following interviewee, in the context of sharing trials data:

> *And that is one potential benefit of sharing, so that then you cut down on the amount of resources that you need to use to get particular results…*
>
> **Interview 13, Statistician and Group Head, African, Research Institution, Africa**

The below interviewee suggested that this may be a motivation for funders to promote data sharing – with a desire to make better use of data collected through their funding streams, rather than it

being hidden away. Technological advances were noted to facilitate the use of data by others, as it was now more feasible to share data online via cloud storage:

> *I mean, the proximate cause is, the proximate people driving it, are funders and journals, in particular [journal]. More fundamentally why are they doing it, yeah, you know, I don't know.* ***I think ultimately people are thinking well, all this research being funded we don't want the data to end up just with people and their local copies squirrelled away.*** *It's more technically feasible now to have an alternative to that. And also, I think people, with all the kind of things that are available in the cloud, then people are thinking well why shouldn't research data be available in the cloud as well.*
>
> **Interview 20, Epidemiologist Researcher, European, Research Institution, South America**

While interviewees sometimes felt that funders may be motivated to promote data sharing to improve the return on the initial investment, developing data sharing platforms has a cost in itself, as discussed above.

In addition, the wider remit of funders to provide grants for research was also considered by some to be structured in a way that it created a barrier to sharing data. The following interviewee explained that as scientists were expected to obtain external funding, they needed to make themselves appealing to these funding bodies through ensuring recognition of their work. She suggested that if they share data, they may not get full recognition for their work, which would impact negatively on their ability to get funding, and therefore progress in their careers:

> *The nature and structure of scientific funding and the financial support of academic, work academic institutions. Plus the nature of publishing. I think that's a smaller issue. It's an important issue, it's a really big issue but it's an important issue. Okay, so scientists depend on getting grants. There's very little discretionary money from universities per say, and even they depend on their research reputation. So everybody is concerned that if they share very early…they accept* ***the indignity of their work stolen through the personal and emotional loss of recognition. Even if they're altruistic not to care about that they will recognise that they need to be credited for what they've done to be able to continue to work.***
>
> **Interview 18, Doctor and Policy Maker, European, Government Funding Agency, Europe**

This again is at odds with the suggestion above of funders being advocates for data sharing, suggesting their wider remit is perceived by some as acting against the rapid sharing demanded in emergencies

5.2.2.3. Avoiding wasted data and samples

A further reason given for sharing data and samples was to reduce waste in research. This interviewee suggested that, by not making full use of data that have been collected from studies, the research participants' goal to do something beneficial to others through the research would not be fulfilled:

> *There is a lot of wasted data that's sat around and not analysed, And for that reason, it's a good thing to share it. I think you've got a responsibility to use what you collect. Otherwise, you've wasted people's time and any conflict that they've gone through to provide that data has gone for nothing. The reason people take part in studies on the whole, there's been research into why people take part in studies, and it's because they want to do something that's going to benefit other people. So, if it doesn't, then it hasn't followed through.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

The quote makes an alignment between data sharing and ethics, couching the justification for sharing in the language of responsibility, and benefit.

The following interviewee explained that in the country he lived in it would be considered unethical to dispose of samples without using them for more analyses, and that it would be considered wasteful. He contrasted this to other contexts in which he worked, where there may be a requirement to dispose of samples after a certain time. The focus here was on re-using the samples to conduct more scientific studies, although this may not necessarily entail sharing them:

> *In some aspects, I often have the impression it's much more strict, the whole process of what can be done with samples in our collaborative partner institutes. I'm not sure whether that goes for all of the aspects but at least also for example* ***with the example that I gave that samples have to be thrown after five years just to make sure that nobody is abusing them. Well, here [Western Europe], that would be seen as unethical, to throw them away because you're wasting patient material.***
>
> **Interview 4, Pharmacologist Researcher, European, Research Institution, Europe**

Of note, none of the interviewees evoked a moral obligation to share samples with others (in contrast to data), merely in this case not to waste samples and to use them for research.

5.2.2.4. Contested priorities of academia and researchers

A notable aspect of the discussion around academia and researchers was a dissonance between what some considered to be the priorities of (other) researchers with respect to data and sample sharing, compared to those that were discussed by researchers themselves. While some researchers were concerned about losing out on recognition for related research outputs from data or samples they have collected, or suspected other researchers were concerned about this, researchers also described a range of other considerations. This included the responsibility they felt to protect research participants, their desire to obtain consent and comply with ethics processes, and to control sharing in order to prevent misuse of data or samples, as will be explored in more depth in Chapters 6 and 7. The dissonance relates to the degree of emphasis given to the different priorities, where a number of interviewees often considered academic recognition to be the predominant priority of researchers, whilst the priorities and concerns discussed by researchers themselves were more varied. This dissonance is discussed in the following quote, in relation to qualitative research on viral hepatitis:

> ***People are telling me about very sensitive stuff.*** *I'm hearing people telling me about being raped, having their children taken off them, seeing friends die, childhood abuse, let alone activities such as selling and buying and using drugs, you know. These are very sensitive, vulnerable people and the subjects are very sensitive and, you know, so that's why I guess I do feel very protective.* ***I feel protective of them and I feel protective of the data and I feel quite strongly that if somebody else is going to come along and interpret that and use it, that they should be in touch with me.*** *I don't think that that is a widely recognised or supported view, actually. I think there is a sense of your being too precious or you just want to get in on authorship or you are being ... Whereas, for me, it's more than that.* ***I think it is really complex and tricky and I think the best thing we can do is just keep on having these conversations instead of saying, no, no, no to data sharing, say, these are our concerns, how can they be addressed?***
>
> **Interview 27, Sociologist Researcher, Oceanian, Research Institution, Europe**

The quote shows the interviewee's discomfort at the assumption that any resistance to share is assumed to be related to academic recognition, and the need for conversations to work through other concerns.

5.2.3. Health care and academic organisations: Protecting patients and research participants

Health care organisations including NGOs, and academic organisations, and their constituent professions were seen as prioritising, to varying extents, the protection of individuals from whom data and samples were collected. While academic organisations were seen as also having a clear priority to facilitate research, while protecting research participants, NGOs in particular were seen as organisations that had an explicit aim of protecting their beneficiaries. They were perceived to have less interest and experience in conducting or facilitating research through sharing data and samples (although there are exceptions to this). As noted above, in the case of infectious disease responses, interviewees discussed NGOs as health care providers, although they also do have other roles. The following interviewee, based in an NGO, discussed the priority given to protecting their beneficiaries, saying that they would not share data where they perceived it would put sections of the population at risk – in this case with governments - if the data related to a marginalised group:

> *We sometimes have to protect our beneficiary populations in Africa from their governments who are not treating them well, because they are a marginalised population, or because they're in a conflict situation, or because there are political disagreements or religious stuff, whatever. You have it now in Europe as well this thing. We need to protect the migrants from the government.*
>
> **Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe**

These protections were often put forward as a tension with the public health priority to share rapidly and early, given that protecting the individuals that data and samples were collected from requires constraints on sharing data. This is illustrated in the following quote, suggesting that

NGOs were torn between their priority to protect people and their wider humanitarian remit, which would potentially align with public health priorities to share rapidly:

> *Then there's the NGOs, who have a humanitarian remit and who are there to just benefit the health of the patients that they treat. So that's a bit of a double edged sword because groups like [international NGO] are incredibly good at protecting their patients and* ***they see patient protection as a very high level priority*** *and they have ways that they as an organisation interpret that within ethical frameworks, within privacy requirements, etcetera.* ***But they also have that humanitarian, benefit the world, improve health remit which can conflict when it comes to data sharing because then the idea is get the data out there and promote the sharing of it.***
>
> **Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe**

5.2.4. Reputation and economy of the country

A small number interviewees discussed the priority of governments to protect the reputation and economy of their country. This led to a discussion about the reluctance that nations may have about sharing data related to an outbreak in their territory, as it may deter travel and trade to the country, thereby impacting on the economy. The interviewee below discussed the issue of a lack of reporting of outbreaks from countries, in this case in relation to Cholera, related to these concerns, suggesting that this was a 'bad' reason not to share data:

> *Emerging infections or epidemics context, I think it's a lot more politically charged and sensitive because use of this data might have some consequences for the authorities which are supposedly in charge of that; consequences that they don't want to deal with. Either, for instance, until the recent update of the International Health Regulations,* ***Cholera, for instance, was a compulsory reporting disease and most of the country which had Cholera outbreak did not report Cholera because they felt it was damaging for their reputation and perception and tourism and all sort of a bad reason not to report it but that was their perception.***
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

A further concern was how benefits would be distributed from research conducted on the shared data and samples, as highlighted by the case of Indonesia withholding influenza samples in 2006/7 (see Chapter 2 Box 2.1).

5.2.5. Commercial priorities

Those interviewees discussing commercial organisations (generally pharmaceutical companies) considered their main priorities to be to develop innovations, and to make a profit.[14] There was a prevailing sense of suspicion around groups who may make use of data or samples with commercial gains from the end products. A concern was that commercial organisation would claim intellectual property rights on innovations derived through sharing, that would then increase the costs of these innovations, putting them out of reach of LMICs. Some did note that the products developed may themselves be important for the public health response and clinical care, and so the concern was the distribution of these benefits. The distinction around whether sharing with commercial organisations was ethical seemed to be in the motive – whether it appeared to be primarily related to seeking profit, or to develop a useful innovation, as discussed in the following quote:

> *…[I am] sometimes approached to provide samples of data to a commercial company, if they're doing that purely for the method of profit and not because it's something that actually you need I wouldn't be very interested…If it's a commercial company who are developing something that nobody's got that is needed then we might well say, "Yes with the right caveats around it." Because it might be something that market desperately needs, a rapid test for example for something and nobody's doing it, then maybe that would be reasonable.*
>
> **Interview 12, Doctor and Laboratory Director, European, Government Agency, Europe**

An issue that was raised with respect to these two commercial priorities of developing innovations that could have public health benefit, and deriving a profit, was that they may be misaligned:

> *The most important research might be the one that does not necessarily lead to commercial stuff. Therefore, the biobank should even be more independent than the data sharing platform, because they need to be sure that there is no conflict of interest, and that samples are not systematically going to entities that might make commercial use out of it later on.*

[14] No one from a commercial organisation was interviewed in this study, so these are the perspectives of those from other types of organisations.

Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe

Relating back to the analysis in Chapter 4, the interviewee makes a distinction between sample sharing and data sharing, because of the scope to develop innovations from the use of samples. She also discussed the importance of independence in decision-making on who can access the samples, which will be re-visited in Chapter 7.

5.3. Priorities and low- and middle-income country settings

Interviewees noted that tensions resulting from differing and sometimes conflicting priorities between organisations and professions were potentially more acute when data and sharing took place from LMICs to HICs. Some of the tensions were practical in nature – related to the challenges in collecting, storing and sharing data and samples in a context with fewer resources. Other tensions were symbolic in nature - that based on a backdrop of colonialism and historical (and ongoing) global injustice, the sharing of data and samples from LMICs to HICs was reminiscent of these wider injustices. Both the practical and symbolic nature of the tensions related to sharing from LMICs were emphasised by the following interviewee, based in a low-income country. He framed sample sharing as a potential way for samples to be taken from low income countries for the development of innovations that make money for those in HICs:

> *And resource rich institutions, resource developed countries have to realise that these resource poor environment, the responsibility is not to provide samples for you to go and analyse, develop assays, and get rich off those assays.* ***You have social responsibility to these resource limited environment…these vulnerable populations.***
>
> **Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa**

The interviewee strongly described an inequity in the sharing relationship between HICs and LMICs, and a responsibility for those in HICs to not use shared samples in a way that was about financial gain. This also alludes to another issue – that commercial organisations such as pharmaceutical companies are often based in HICs.

The disparity in resources was also discussed as making the concern around academic recognition particularly acute, where data and samples were shared from LMICs. This interviewee, both from and based in a LMIC, highlighted the disparity in resources available to analyse data:

> *…**sometimes I don't know that its fear ... yeah, sometimes it's the fear actually that "I have done this work in a resource limited setting" then, what happens is that somebody else who has much more powerful tools can actually do much more work with that data and identify things, of course, which I could not have seen in my limited, you know, setting or surrounding.** Sometimes ... what I'll have, sometimes people say, the people at the receiving end, they're receiving the data. They're not doing all this work, they're doing it in the background, you know, so they have more time to sit down, publish more papers and you never know what it is in the end, the ending game of it. You probably could have published one paper but they end up publishing ten papers, that kind of thing.*
>
> **Interview 5, Clinical Trials Manager, African, Research Institution, Africa**

The interviewee clearly expresses concern that while those in the LMIC may have put in a lot of work collecting data, if shared, it would then go on to be analysed by those with more resources in HICs who may then get more publications and therefore more credit for the work.

The issue outlined above of academic recognition in relation to data collected and shared from LMICs was echoed by the interviewee below, both from and based in a HIC setting:

> *The importance of the career structure and in really classical frameworks of value-adding that career structure and by that I mean the importance of first and last author publications. Very strictly. The importance of the number of publications and just ensuring ... It's like an accelerated or amplified version of the academic competition and hierarchy that is here. It's even more intense, I find, in low and middle income countries. And then additionally the history of North/South partnerships starting with colonialization and moving on to scientific partnerships as they exist today, **I think has engendered protectionism…among some low to middle income country researchers.** And so put **all those things together and data sharing is a very difficult concept to grapple with because it means that people are exposing themselves to enable more science that they cannot necessarily take part in or benefit from. If they don't get credit for the***

> ***data that they've shared, if they don't have the resources required to enable their own use of shared data that actually has potential of increasing the research gap between North and South academic institutions.***
>
> **Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe**

The quote shows concern from the interviewee that the gap in research outputs between academic organisations in HICs and LMICs (referred to as 'North' and 'South' respectively) could potentially increase as a result of sharing. She also suggested that the relationships between HICs and LMICs both in relation to colonialism, and more specifically in current experiences of scientific partnerships, have led those based in LMICs to be more protective of collected data.

Given the disparity in resources between HICs and LMICs, capacity building was viewed by many interviewees as an important obligation for organisations involved in data and sample sharing. One interviewee, involved in the Ebola response, discussed the development of an infectious disease biobank. He suggested that the development of a biobank of samples that had been collected during the Ebola outbreak would provide a focal point to train researchers, and foster collaborations both within the region and beyond:

> *…obviously the aim is to better understand and better characterise the epidemic of [LMIC]. So the intent is to establish a framework called Biobank in [LMIC], and to be able to utilise and to train [West African group] to be able to analyse those samples and foster research collaborations among [West African] scientists as well as collaborations with other scientists and researchers that are interested in a wide range of studies, including genomic as well as immunological, that is antibody response, and as well as testing assays.*
>
> **Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa**

This shows capacity building was not considered to be merely an incidental positive outcome, but was seen as a clear motivator for the biobank development.

Capacity building was not, however, seen as a straightforward solution to the inequities potentially arising when sharing between LMICs and HICs. The following interviewee, who had extensive experience developing data sharing platforms, considered capacity building in relation to data sharing to be potentially problematic. He suggested that the skills developed from sharing data, in this case conducting an individual patient meta-analysis, may not be the priority skills required in a particular country:

> *What I'm really keen on doing is instead of making sure that everybody in the world will be able to do a complex and individual patient meta-analysis as sort of a trade-off of sharing data, it would be just to make sure that we are responding to the demand of the countries and the investigator of their priority and basic demands, as opposed to us creating expectations or demands which are not the priorities.*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

The implication here is that looking at the bigger picture, developing the skills related to analysing shared data may divert people from developing skills that are more of a priority**.**

5.4. Conclusion

The findings presented in this chapter make it clear that there are a number of organisations and professions involved in data and sample sharing, which have different roles in relation to infectious disease prevention and response. The WHO was seen as the lead organisation for the infectious diseases response globally, and therefore noted to be a key organisation with respect to data and sample sharing. It was also seen as a locus of sharing, being directly involved, although at times seen as falling short of expectations.

Funders and journals were both seen as promotors of data sharing, although the top-down nature of this promotion was a source of concern for some interviewees who felt it did not adequately take account of the concerns that those being called upon to undertake sharing may have. Other

concerns focused on practicalities such as providing funding to enable sharing to happen, or guidance on how to sharing ought to be done.

A number of professional priorities emerged in relation to data and sample sharing in the infectious disease context. These priorities did not always map neatly onto organisational typologies. Some priorities were noted to be cross-cutting – for example organisations across the board were considered to have a role in sharing data to facilitate the public health response. Other priorities, such as the pursuit of academic recognition, were perceived to be more specifically the concern of some organisations and professional groups involved with research.

Some priorities appear to be complementary and well aligned, such as public health priorities to share data and gather knowledge, and academic priorities to do more science, maximise funding and reduce waste in research. Others appear to be in conflict. Some priorities that could limit sharing, for example, protecting participants and patients, delaying sharing to gain academic recognition, or avoiding declaring an outbreak in a territory, could be in conflict with the priority of rapid sharing.

Another notable sub-theme is that commercial organisations were sometimes viewed with suspicion by interviewees, who seemed to suggest the profit motive was a 'bad' reason to be involved in data or sample sharing. Interviewees did also consider organisations such as pharmaceutical companies to potentially add value if they are able to use the data or samples to develop important innovations.

The LMIC context lent an additional acuteness to some of the identified tensions – with these settings both having a greater public health need, due to the greater burden of infectious diseases, but also potentially unable to afford or take up the benefits derived from work on shared data and samples. There was also concern that researchers in LMICs could lose out on academic recognition, if data or samples they had collected were analysed rapidly in HICs.

This chapter has focused on the theme of organisational roles and priorities, and some ethical issues involved in this topic were alluded to. The next chapter will move to focus on the perceived ethical responsibilities of individuals engaged in data and sample sharing in the infectious disease context.

6. Meeting responsibilities and maintaining control

The previous chapter showed that numerous organisations and professions are involved in sharing data and samples in the infectious disease context, and that these different organisations were perceived to have differing priorities. At times these priorities are complementary to each other, and at other times are in conflict.

This chapter follows on from the previous by moving from a focus on institutions and professions, to the individuals themselves, and their own decision-making. The chapter explores two interlinking themes emerging from the analysis - the perceived responsibilities of those engaged in sharing data and samples, and the desire for control over data and samples. Interviewees discussed a number of responsibilities they viewed themselves as having to patients, research participants and others across the sharing process, from data and sample collection, to sharing and future uses. Maintaining control of data and samples is a notable related theme, with interviewees citing control as a way to ensure that data and samples are used appropriately, by people with the right skills and for the right purposes. For some interviewees, the overriding moral duty was to share data for the public good, and to make optimal use of samples.

6.1. Responsibilities of those involved in sharing

This section explores the responsibilities that interviewees considered themselves, and others, to bear with respect to data and sample sharing in the infectious disease context. These responsibilities were predominantly discussed in relation to three sets of people: firstly, responsibilities to the people from whom the data or samples were collected (section 6.1.1); secondly, those to the people who may benefit from knowledge and innovations derived from the shared data or samples (section 6.1.2); and thirdly, to other professionals who may be involved in data and sample sharing (section 6.1.3).

6.1.1. Responsibilities to people from whom data and samples were collected

6.1.1.1. Duty of care: recognising the people behind the data and samples

Those with a health care or research background often felt a duty of care towards the people from whom data or samples had been collected, conveying a sense of personal responsibility that data or samples that relate to their research or clinical care are protected. The nature of this protection included issues such as respecting the wishes of the people from whom data and samples had been collected, maintaining privacy, preventing exploitation, and honouring the trust with which individuals had shared information about themselves with the individual researcher or clinician. One interviewee, based in an NGO, explained that in her view the data and samples are a part of the patient. Caring for those data and samples was therefore something she viewed as a part of her wider role in caring for the patient:

> *I don't see much of a difference between taking care of them clinically and making sure that their data are not being spread around the world, were not protected, or being exploited. Or their tissue, most certainly the case, I think, you know, their tissue is still them. And* ***taking care of your patient includes taking care of their tissue and taking care of their data.*** *And I think it's easier to think of the patient as just what's sitting there. When the specimen goes off, and the data go off, and you think that that's your primary locus of caring. But I think that notion of caring needs to be extended to everything. Everything that is that patient, including data and samples.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

When discussing sharing she used the phrase "spread around the world", which suggests a view that sharing data globally could be a careless activity. This interviewee recognised the potential public health benefit of sharing data and samples, her concern was around how the sharing was done.

Distance of data and sample users from the participant/patient

There was also concern that some of those using shared data or samples were more removed from the individuals from whom they were collected both physically but also emotionally, and therefore

they might not have the same level of concern for those individuals. This is explained by the interviewee below, who felt acutely aware of the person behind the data or samples:

> *There are scientists who only work on data. They don't generate their own data. That's their job to work on other data. And* ***they almost have no interest in the people, you know. But actually they need to be aware that these samples ... These data come from people, and there is a chain of events that come up to that. They've got to be mindful of that.*** *Therefore, they should always have to feel they should justify what they're doing in terms of the common good.*
>
> **Interview 15, Doctor and Scientific Director, European, Research Institution, Africa**

The quote above suggests a link to how data and samples go on to be used. The interviewee states a clear ethical obligation that they ought to be used for the "common good" – that is, the data or samples ought to be used only for purposes that fulfil this aim, which appears to be linked to a sense of respect for the people from whom data or samples were collected.

<u>The infectious disease context as a source of vulnerability</u>

The infectious disease context was considered to add additional considerations with respect to the duty of care that interviewees felt towards patients and research participants. In particular, interviewees considered that the people from whom data or samples had been collected were inherently vulnerable because they were affected by an infectious disease. With respect to data and samples collected in emergency contexts such as the Ebola outbreak in West Africa in 2014-16, interviewees considered people vulnerable due to the chaos and fear surrounding the outbreak, the lack of choice they were perceived to have in seeking medical care and having data and samples collected from them due to the particularly severe nature of the disease, and the potential gaps in following due process with respect to good ethical practice for collecting data and samples for onward sharing:

> *I think it's a complex issue, complex issue in a sense that when samples were collected from these participants, I mean for these at risk, vulnerable you can say individuals or participants, or whatever you want to classify them, there were nothing in the stipulation - like I say, there was a crisis situation,*

> ***everyone was desperate to ensure that the crisis come to an end, so there was a lot of gaps, issues that were not addressed during the emergency. It's an emergency, almost like, for example, your house is on fire.***
>
> **Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa**

People from whom data and samples had been collected outside of emergency contexts were also considered vulnerable by some interviewees. With respect to diseases such as Visceral Leishmaniasis (VL), research participants were considered vulnerable by interviewees because it was seen as a disease that more commonly affects poorer people. As a result the people from whom data or samples had been collected were considered to be at particular risk of being taken advantage of, and in particular need of protection:

> *I think one of the key things is that the areas that we work in affect the poorest, or the poor populations. As such, they're considered very highly vulnerable and prone to ... maybe let's call abuse for now. That* ***you can exploit or take advantage of the situation if we don't put any measures to protect them.*** *So, there's high levels of illiteracy and especially in research, might not be, what do you call it ... they are willing and give consent but the understanding may not be totally ... how do I put it. Let me use the word hundred percent but what I actually mean is that they're not aware of all the issues regarding research, which of course, technically you'd have much better information and to know what can happen potentially to those, whether they are samples, whether it's just the data, and all this.*
>
> **Interview 5, Clinical Trials Manager, African, Research Institution, Africa**

Along with others, this interviewee discussed this explicitly as exploitation, highlighting concerns about fairness, which are explored further below.

A further vulnerability was considered to stem from the marginalised and criminalised nature of groups who more commonly suffer from certain infectious diseases, and the associated stigma of having this disease. This was considered to lead to a risk that the data may be analysed, interpreted and presented in a way that was denigrating towards the participants in the study. This was highlighted by the following interviewee, who researched groups with blood borne viruses, and

considered the population she worked with to be already marginalised by society. She was worried that shared data would be analysed and presented in a way that was further stigmatising:

> *You know, as I said, I think people who inject drugs are one of the last sectors of society in which it seemed to be okay to stigmatise and marginalise.* ***A lot of the people that I happen to interview have come from history of trauma. They don't have socially validated lives.*** *They may not be in full employment. There are often dependent on illicit substances. They often have to engage in the black economy to fund their habits. All sorts of things like that. So it's a whole sort of world that's going on but which is often quite denigrated in our society. For that reason, they're vulnerable and they're also vulnerable because of those things, they're vulnerable to state sanctions, criminal sanctions and actions like having their children taken off of them and being imprisoned. All these things. A lot of people have experienced those things and have been deeply traumatised by those.* ***So, you know, I have a responsibility, I feel, to treat my participants…any participant, with respect. And be respectful of how I interpret what they tell me and who gets to see it.***
>
> **Interview 27, Sociologist Researcher, Oceanian, Research Institution, Europe**

This quote also points to a lack of trust in some of those who may use shared data, trusting that the data will be analysed and presented in a way that remains respectful to the original study participants.

These questions around vulnerability and associated duty of care were noted by some interviewees to be a part of good research practice in general, irrespective of onwards sharing. This is raised by the following interviewee in the quote below, who felt that the issue of vulnerability and duty of care was one that resided at the point of data collection:

> *The more vulnerable a person is, it's more of a responsibility you have for the participant. That ought to apply at the point of data collection. By the time you get to data sharing, does it make a difference? Unclear.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

6.1.1.2. Responsibilities to maintain privacy and confidentiality

This next section will focus on another facet of the perceived responsibilities towards the people that data and samples had been collected from – the safeguarding of privacy and confidentiality. The concern was considered particularly acute for data that may be about a small number of

individuals, for example in a small outbreak, or where a number of datasets were going to be linked and where people may, for that reason, have the potential to be re-identified. Interviewees were particularly concerned that if people to whom the data relate were identified, it could lead to potential harms to that person – that they would be revealed to be infected with a stigmatising disease, or that with respect to certain infections, they could be criminalised. The following interviewee discussed this in the context of HIV:

> *Particularly on a global level, there are sort of laws which pertains to this, like different countries which could criminalise an act of infecting somebody with HIV. And if you have evidence that could uncover that, then of course, that has again more reason why* ***you've got to be very, very careful that our data are stored and provided in a form that doesn't allow individuals to be identified who were either infected or, you know, can identify who has infected them.***
>
> **Interview 15, Doctor and Scientific Director, European, Research Institution, Africa**

This sharpens the dilemma faced by people who felt there was a risk of potentially severe consequences for individuals that the data relate to, while at the same time there may be real benefits for sharing for public health purposes – for developing responses. It also highlights the particular tension with respect to infectious diseases, which cross borders between different legal jurisdictions and where the law may be very different with respect to criminalisation around disease transmission and behaviours that may be associated with the disease.[15]

A challenge to meeting this responsibility was the ability to truly anonymise data, with a number of interviewees expressing a lack of clarity around standards of anonymisation, as discussed in Chapter 4. The interviewee below, who had expressed a great sense of responsibility towards those participating in her qualitative research, was concerned that the data could not easily be meaningfully anonymised, as people involved in her research were discussing personal details

[15] For more information on the criminalisation of HIV transmission see: Barré-Sinoussi, F., Abdool Karim, S. S., Albert, J., Bekker, L.-G., Beyrer, C., Cahn, P., Calmy, A., Grinsztejn, B., Grulich, A., Kamarulzaman, A., Kumarasamy, N., Loutfy, M. R., El Filali, K. M., Mboup, S., Montaner, J. S. G., Munderi, P., Pokrovsky, V., Vandamme, A.-M., Young, B. and Godfrey-Faussett, P. Expert consensus statement on the science of HIV in the context of criminal law. J Int AIDS Soc. 2018; 21(7):e25161

about their life that could enable someone to identify them. This shows the worries around whether anonymisation will work, that is whether removing identifiers renders something anonymised, and what risks remain:

> *There's so much, a wealth of data. People stories, their family makeup, they're talking about their schools, they're talking about their trajectories that you can't guarantee that somebody won't be able to piece together who that person is.*
>
> **Interview 27, Sociologist Researcher, Oceanian, Research Institution, Europe**

A minority of interviewees did not consider issues of privacy and confidentiality to be a prominent source of concern. The interviewee below discussed her own experience of the relevance of international requirements, guidance and norms to individual contexts where people may have different norms and values with respect to privacy and confidentiality. She went on to explain that she considered confidentiality to be something that was prioritised in 'Western' cultures, but that in the settings she had worked in in South East Asia, people had a 'community approach' which meant individual privacy wasn't considered so important or prized:

> *Because Eleanor Roosevelt chaired the committee post second World War that sat down and decided what the human rights treaty or policy, whatever it's called, would engender. And that was the rights of the individual because it was all American. It was all Western people who wrote that and so this has been proliferated amongst, I think, normative ethical approaches to this type of interaction in that we have to protect the individual, protect the individual. Now, having worked for the majority of my career in [South East Asia], I would say actually that's inapplicable to many settings.* ***In [South East Asia], for example, it's very much a community approach to culture, to protections, to thinking, to how people live their lives. So it's the community that requires protection, not individuals and things are very open and very shared****.*
>
> **Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe**

This also highlights another particular issue that individuals may face – adhering to international norms, which may not be complementary to working in a way that is in keeping with the norms of the culture they are working in.

6.1.1.3. Responsibilities to share benefits

A notable area of discussion from this analysis is the distribution of benefits resulting from improved knowledge and innovations. These benefits can be varied in nature, and relate to both knowledge and innovations aimed at preventing disease and interrupting its transmission, through to treatment and cures.

Interviewees saw a fair distribution of benefits as a part of responsible data and sample sharing, and were concerned to avoid being involved in an activity in which the benefits were not fairly distributed. The level of concern differed by disease type, and was more acute for those diseases that are more geographically widespread and are perceived to pose a potential threat to high-income countries (HIC) as well as to low- and middle-income countries (LMIC), such as Ebola. For diseases that also pose a threat to HICs interviewees were concerned they could be sharing data or samples in a way that results in an unfair distribution of benefits, due to different resources between HICs and LMICs to benefit from improved knowledge of prevention and treatment. This was heightened further where there was potential for innovations to be developed, particularly from the sharing of samples, that could lead to intellectual property claims by those who go on to use the shared samples. This concern was voiced by the following interviewee:

> *…while [sub-Saharan African country] is donating blood for that research, maybe one in the North, after using this blood for research and coming out with another produce,* ***if the benefit of that product remains in the North, not coming to the South where the research has been done it's clearly injustice****, according to ethical principles.*
>
> **Interview 9, Pharmacologist and Principal Investigator, African, Research Institution, Africa**

The quote also shows that this was discussed explicitly in terms of ethics.

Different interviewees had different views about what was entailed by a fair distribution of benefits. At times it focused on the benefits coming back to individuals or communities that data

or samples had initially been collected from, whilst at other times it was seen in a global public health context, with fairness consisting of knowledge and innovations being used where they are needed most, and where the burden of disease is.[16] This interviewee explicitly noted this issue that the search for global public health goods does not necessarily lead to benefits for particular individuals or countries:

> *When something becomes a global threat, then it attracts everyone's interest and everybody, rightfully, has some interest in it. And every good ... not going [to] say every good ... but if I limit myself to drugs, vaccines and diagnostics. So every drug, every vaccine, every diagnostic that could come out of research using these samples could potentially ... and again, potentially ... be beneficial to anyone and everyone in the world.* ***So in that respect, when you think about it, you would say that every effort should be made for these samples to be made available to anybody and everybody who is willing to do work that will result in a global public health good. But global public health goods come along with limited benefits to individuals, to corporations, to countries, and this is where it becomes tricky.***
>
> **Interview 28, Doctor and Scientific Director, African, Research Institution, Africa**

Although the predominant focus was on the wider community or country in which data had been collected, benefits back to the specific individuals from whom data or samples were collected were also discussed by some interviewees. The following interviewee raised a specific issue related to certain infectious diseases – that there is anticipated immunity from the disease (for a period of time at least) that may mean that knowledge arising from sharing is not of direct benefit to that individual. This was a source of concern for her, and she went on to discuss that she was unclear about who she ought to be prioritising – whether she should be more concerned with her perceived responsibility to the individual from whom the data had been collected, or whether her priority should be wider public health. In this case the worry is heightened as the individual is not likely to gain direct benefit from the sharing due to the nature of the disease:

[16] Of note, with certain infectious diseases the answer to these two questions may not be the same, particularly for diseases that do not have stable ongoing transmission in certain locations. In the case of Ebola at the time of writing, the 2013-16 epidemic in Sierra Leone, Liberia and Guinea - often discussed by interviewees in this study - was over, but there was a subsequent epidemic in the Democratic Republic of Congo.

> ***So we're talking about data from vulnerable people. And that's a concern as well, you know? Who benefits from this? Not them.*** *Unless you imagine that somewhere down the line they're going to benefit. Certainly the case, interestingly, with Ebola, that if you're doing an Ebola survivor study, those people can be expected not to benefit. They've had Ebola. They're not likely to get it again. Well, we don't know, but we would expect if it's anything like Marburg, you probably have ten to fifteen years of resistance to reinfection. So they will not directly benefit.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

The people from whom the data were collected were again characterised as vulnerable here.

6.1.2. Responsibility to the wider population

Another group of people that many interviewees considered themselves to have a responsibility towards was the wider population. The wider population was either other people (who were not the people from whom the data and samples had been collected) who at that time may benefit from knowledge and innovations from data and sample sharing, both within emergency and non-emergency contexts, or people who in the future may do so, for example during a future outbreak. This is closely linked to the analysis of public health priorities in Chapter 5 to save lives and improve care.

Some interviewees went on to frame this responsibility to the wider population as leading to a moral obligation or imperative to share data (it was not mentioned for samples), because of the potential public health benefits. This is illustrated by the quote from the interviewee below:

> *But then again it comes back to well, is that data you hold though, could that be used for public good? And then, for me, it's more a moral imperative.*
>
> **Interview 22, Doctor and Group Head, European, Research Institution, Europe**

The reasons invoked for this moral imperative are consequentialist in nature. This was also put forward as a professional responsibility, as discussed by this interviewee:

> *I think that most people who are doing this kind of work are doing it because they want to improve knowledge, and or the quality of life for whoever. Or they want to improve the survival of the planet, whatever. So, and to me there is a moral imperative for scientists, doctors, public health people to try and make things better. So, yes it's very simple minded. I think it is a moral imperative.*
>
> **Interview 18, Doctor and Adviser, European, Government Funding Agency, Europe**

The responsibility to the wider population was felt by some to be in tension with that to the individuals from whom data had been collected. In this quote, the interviewee describes being torn between these competing interests:

> *I just don't see a model for acting in anyone's best interest. And everybody resorts to the, "Sharing is good. Collectively, everybody's going to benefit because we're going to do all this research."* ***And it's like, well, there's always an ethics trade-off between public health ethics, interest of everybody's health, and individual ethics. And no one can sit in the middle of that and tell me what to do. And until that happens, I'm not going to feel better about this.***
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

The quote above shows that this was a decision that weighed on her personally.

It is also notable that sharing rapidly and widely was not always seen as a straightforward way to meet responsibilities to the wider population, due to concerns that data may not have been quality assured, or the data or samples may be used for bad quality research, which could have implications for how infectious diseases are prevented or managed, as noted by this interviewee:

> *There is risk of misinterpretation, so you need to ... and it's data which,* ***if it's misinterpreted, could actually lead to unnecessary damage, economic damage to the country, unnecessary worry to the population****. The concern about misinterpretation and misinformation using the data is the challenge. Data misused by media or organisations can actually be harmful, so it's how that data ... So the open data is critical, but this is the core data which ... and the other thing about the core data, is it can contain lots of mistakes, and that was recognised in Ebola, so actually there was lots of duplication, it wasn't cleaned. A lot of mistakes can be made by using the core data. It was that balance of [what] do you make of uncleaned data.*
>
> **Interview 14, Doctor and Policy Maker, European, Government Funding Agency, Europe**

Another challenge that was highlighted is that the wider population who may benefit from the knowledge and innovations developed as a result of data and sample sharing is not homogenous. There was concern in particular that those in LMICs would not benefit as much as those in HICs from the knowledge and innovations derived from data and sample sharing, as discussed in section 6.1.1.3 above, and in Chapter 5.

6.1.3. Responsibilities to other professionals involved in sharing

The responsibility of giving acknowledgement and recognition for the contribution that fellow professionals have made in the data sharing process was often discussed. This was considered more acute where sharing was taking place from LMICs, with onward use for research by people in HICs, due to the discrepancy in available resources. The following interviewee discussed the need to attach 'kudos' to acknowledgement given for data sharing:

> *…when data is used for a secondary analysis then I believe that all of the primary authors of that data where it originated, or what wish to be, should be acknowledged within the publication and so should be aware and that's part of that academic credit but then also that that acknowledgement should have some sort of kudos, ideally, and some metric which is of relevance to their careers that can be evaluated.*
>
> **Interview 17, Clinical Trials Researcher, North American, Research Institution, Europe**

As noted in Chapter 5, acknowledgement and recognition were considered to be particularly pertinent to people working in academia. Varied ways of acknowledging others were discussed, from authorship on scientific papers to mentioning the contribution of others to data or sample collection. Some interviewees advocated for a special system of recognition for data sharing within universities.

The risk of losing recognition for work undertaken was noted to be a real concern for those who may contribute data and samples for sharing. Without this responsibility being fulfilled, those who contribute their data may feel used:

> *One is that you'd find that there's a fear that the local researchers have that, when they share data then, other people take all the credit for the work over them. That* ***then they end up being, if you like, just tools, for lack of a better word.*** *So they end up appearing like spanner boys.*
>
> **Interview 13, Statistician and Group Head, African, Research Institution, Africa**

6.2. Maintaining control

As discussed above, interviewees considered themselves to have certain responsibilities with respect to data and sample sharing, particularly towards the people and wider communities from whom data or samples had been collected. Interviewees also discussed particular rights they had to control data and samples that they were involved in collecting or curating. This was often (although not always) linked to interviewees' perceived responsibilities – that they needed to control what happened to the data or samples to ensure they met their responsibilities.

This section starts by discussing what interviewees discussed as being in control, who they considered ought to have control, how it manifested, and why control was considered important. This raises a number of questions on the ethics of data sharing – such as whether ownership rights can be claimed over data and samples in the context of an infectious disease response, and who ought to make decisions about whether, and with whom, data and samples are shared.

What control meant ranged from a claim that data or samples belonged to a specific group of people, to a weaker claim that certain groups of people, often those who had contributed the data and samples, should be involved in decisions about what they are used for, and who they are used by.

6.2.1. Manifestations of control

6.2.1.1. Ownership as a form of control

Ownership was raised by a number of interviewees, particularly those based in sub-Saharan Africa. For samples, ownership was more generally claimed to be that of the patient or participant, or the

government of the state they were collected in, with those involved in collecting them becoming gatekeepers with respect to their use. Many interviewees, however, discussed an uncertainty with respect to ownership of data – with the issue of who owns the data being raised but not resolved for them. Some interviewees, such as the one quoted below, suggested that without recognition of the ownership of the data by the researcher the activity was data taking rather than data sharing:

> *You have obtained [the data] from the clinical trials, which you did. Then, if your data are going to be taken by some other people, and these submit they are their own that means violating one's own data rights.* ***So what I think is, those who are for data sharing, they should be sure that there is a mechanism by which one's own data are going to be preserved. It should not be taken by other people.***
>
> **Interview 9, Pharmacologist and Principal Investigator, African, Research Institution, Africa**

The same interviewee also drew an analogy between copyrights for books and ownership of data generated from research activities, suggesting that ownership of data should be retained by the researcher who conducted the study from which data are shared:

> *When we think of data sharing, there must be mechanism of preserving one's own data, of saying these data are my own, not others. Prior to getting, you know … A good example is copyrights. When we write a book, for example, if someone copies your book and you realise it, it is violating the copyright.*
>
> **Interview 9, Pharmacologist and Principal Investigator, African, Research Institution, Africa**

In some cases, more notable with samples than data, ownership was considered by some interviewees to reside with the community or state that they had originally come from. This was particularly discussed in the context of Ebola samples from the 2014-16 outbreak in West Africa, as illustrated in the quote below:

> *But those samples already become [the Government's] property, and we have - and before anything is done on those samples, we have to be aware, and we have to give the approval, and we also have to be a part of whatever study that is being conducted.*
>
> **Interview 21, Public Health Doctor and Researcher, African, Governmental Agency, Africa**

This link between ownership and control is implicit here, with ownership enabling decision-making and involvement in subsequent work on the samples. The relationship between ownership and control is not straightforward, however. Ownership in some instances may be retained by one party, while control resides with another party. This is more in keeping with the model discussed by the interviewee below who considered ownership to reside with the patient from whom the sample (in this example) was initially collected, rather than the researcher – although it is the researcher or ethics committee who will be in a position to control what happens with the data and samples. The interviewee described his own discomfort that some collaborators discussed the samples, and by extension the patient, as their own:

> *The patient is of course, owner of the data... the sample and we ask permission to use that sample for a specific purpose either directly or indirectly. We ask permission through ethics board or something similar. We ask permission through a consent statement but it's a bit different, how do I say that, interpretations of this. For example,* ***we have collaborators that work clinically so they are the ones that are taking the sample from the patient and then they automatically talk about the patients as being their patients, almost owning the patients but also owning the samples that come from these patients which I always find a bit troublesome to deal with.*** *Also because you are, first of all, neglecting the pride of the patient, their own ... They are not owned by anybody.*
>
> **Interview 4, Pharmacologist and Researcher, European, Research Institution, Europe**

6.2.1.2. Gatekeeping of access to data and samples as a form of control

A further mechanism of control discussed was by becoming a gatekeeper. The data or sample contributors – that is the researchers who initially collected the data or samples – were seen by a number of interviewees, particularly researchers, as the gatekeepers for whether or not data or samples ought to be shared, and what can be done with them. This is in contrast to the policy direction by some funders and journals mandating data sharing from research that they have funded or publish.

The interviewee, a statistician based in Africa, discussed his experience when dealing with requests to share data, and the need for permission from the principal investigator, who in this case becomes a gatekeeper or controller for access to the data:

> *If I remember that experience with the one, then when it got to sending the data, I remember the person was communicating to me kept asking me to send the data.* ***I was like, no. I can't send the data unless the investigator tells me so.*** *I remember even being sent for a document to sign up. Then I was like, no, it's not for me to sign the letter committing to give all the data. It's the investigator who needs to commit to that.*
>
> **Interview 13, Statistician and Group Head, African, Research Institution, Africa**

In this quote, as with some others shown previously, the word 'give' is used in association with data sharing.

Another way of assuming the gatekeeping function was through restricting how much of the data is shared. This is illustrated by the following interviewee, who discussed her experience with a project that involved data sharing, and the reticence of the team to share the whole data set. She suggested that the team may have wanted to retain control of the data set, and so would only share the parts that were relevant to the research question being posed:

> *It was a collaborative project and they [group based in HIC] helped with money, with ideas and all that. But locals were saying this is our data. If you have a specific research question, you raise it and we give you that data. We're not going to give you … They wanted 100% control over the data as much as the [anonymised] team was having control over it. Now, don't ask me the reason why they said no but the point was control, that's the impression I got.*
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**

By doing this the researchers appear to be attempting to reconcile sharing and control - they can be said to have undertaken sharing, while at the same time retaining control.

Gatekeeping was also discussed with respect to sample sharing, as illustrated by this quote from a principal investigator who discussed his feeling that once sample sharing has taken place, if the

team who received the samples would like to do other testing beyond that for which the samples were specifically sent, they have to refer back to the team that sent them:

> *So I think for other things we have to know that where the sample are going to be stored and who is going to the sample and I have to know if specifically also ... For example we sent samples for some tests, so these samples should not be used for others. Otherwise I mean they have to refer back to us. They have to refer back to us if they want to do anything else.*
>
> **Interview 11, Doctor and Principal Investigator, African, Research Institution, Africa**

This again puts the researchers who contributed the samples in a position of control.

6.2.2. Reasons for seeking control

There were three main reasons given to seek control of data and samples. Firstly, and primarily, to meet the responsibilities to others that were outlined above in section 6.1. A second reason was to maintain ties with the data and samples. Finally, a third reason was to avoid negative assessments of the data quality, by others who may go on to analyse them. Each of these reasons is explored in the following sub-sections.

6.2.2.1. Control to meet responsibilities

As noted above, the desire for control, perhaps through ongoing recognition of ownership, or some involvement in decision-making for what happens to shared data and samples, was often grounded in concern around meeting perceived ongoing responsibilities – most often in relation to the people that the data or samples had been collected from. The following interviewee discussed her sense of ongoing responsibility for the data after they have been shared, and her worry about losing control about what actually happens to them once they have been shared with a repository:

> *I also think there's a moral argument for clarifying this [duplicating copies of data]. You really shouldn't say, "All right, we're going to take care of these data. Right, this is my data, I'm going to take care of this. But I'm going to make a copy and put it over here, and I then make a copy and put it over here. And then we'll just forget about those." Again, where does my responsibility stop? Am I just responsible for this, and*

> *then I've just given this to you, and I don't know under what conditions those are being kept, or whether or not I could even make a copy? I think that's my responsibility, still.* ***I don't think that organisations who the individuals ask for these data, or the organisation who has set up the platform for asking, or for managing these data, are actually considering that I might still have an obligation here.***
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

This quote again highlights the very personal sense within which researchers see their responsibilities – that they feel directly responsible for what happens, and in some cases this leads to a sense of ownership of the data and/or desire to control how they are used:

Some discussions on meeting responsibilities both to the people from whom data and samples had been collected, and the wider population that may benefit from the sharing of data and samples, suggested a mistrust of some of the people who would go on to use the shared data or samples, and concern that they would misuse them. Discussion of misuse was tied to responsibilities that interviewees considered themselves to have – to prevent bad research, including that which may be denigrating towards the groups to whom the data or samples relate, or that may result in poor quality research that have negative impacts on the wider population, where these inaccurate results may be applied. Misuse was also characterised as using data or samples without due credit to the contributors, or using them primarily for financial gain. Retaining control and restricting sharing was seen as one way of preventing these types of perceived misuse:

> *Restricted access [is important] so that it is not misused. And all data should be used for the good, for the patient good, in my view. I mean, any information that is generated out of any data set, I think should be, in a way, having some impact on the patient, in terms of care, in terms of treatments that are being put in place.*
>
> **Interview 8, Clinical Trials Manager, African, Research Institution, Africa**

Where data or sample misuse was discussed, many interviewees had difficulty naming examples where this had occurred, but it was something they had either read about, or had heard being

discussed. There was also uncertainty about the potential impact of misuse, as highlighted by the quote below:

> *I just hear people talking about [abuse] and they say, "Oh data can be … When you have a data set like that and you easily give it to anybody, people can use it to cause harm." But I've never understood what kind of harm that would be.*
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**

This suggests that this concern is not necessarily grounded in first-hand experiences of misuse having happened for this set of interviewees, but a concern that stems from the potential that it could happen.

A further responsibility discussed above was to other professionals. Some interviewees linked the control of data sharing to the idea that people in LMICs were at risk of being taken advantage of, reminiscent of the discussions in Chapter 5. Control was linked to mitigating this risk. This interviewee was involved in developing a data sharing platform. He discussed the concern generated among investigators that the data would be hosted in Europe – far from the place where the data were collected and where the burden of disease is, and resulting in unfair collaborations:

> *Any issue I think, not a big issue but people would always want to ask, especially if the data sharing platform is probably hosted in Europe, and most of this data is coming from developing countries, so people will always want to raise concerns.* ***Why Europe? Why not hosted somewhere in Africa,*** *I mean Africa is synonymous with developing countries. So, for example, the [disease] data sharing platform is hosted in [HIC-based university], but most of this data is going to come from India, and also Africa, so the question is why [HIC-based university]? There is not any [disease] in the UK, why host it in [HIC-based university]?* ***Some investigators would think that these people, they want to take advantage of us, they want to take our data, and use it in [HIC-based university], they don't generate any data there.***
>
> **Interview 19, Data Manager, African, Research Institution, Africa**

In this quote, controlling data was linked to where the data sharing platform is located, with a clear indication among researchers he had encountered that they would like the data to be hosted in Africa.

The findings above suggest that concerns about not meeting their responsibilities, or worries about others not meeting their responsibilities, led some interviewees to want to constrain data or sample sharing. One interviewee, however, cast doubt on the idea that people were not sharing data because they were concerned about the ethical issues, suggesting that ethics was being used as an excuse because they did not want to share anyway:

> *Ethics…can be used as an excuse for not sharing data. I think people either by overusing or over-stretching what is the reason why ethical concerns should be looked at when you share your data would use ethics as a reason why not sharing data...As opposed to, they have not said, but did not want to share the data.*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

This quote shows the perception that 'ethics' may be used as an instrumental tool to meet other objectives, for example not sharing because of concerns around negative judgements about the data quality (explored more in section 6.2.2.3).

6.2.2.2. Control to maintain ties with the data and samples

One source of concern for interviewees who were based in sub-Saharan Africa in particular, which was linked to a desire to control sharing, was becoming disconnected from the data or samples. Data were discussed as if they are a physical object being moved or transferred across to different places, and becoming more distant from the contributor. This was also linked to ownership of the data and samples, and a concern that through sharing the data or samples the ownership is being transferred or violated in some way. The interviewee below discussed the sense that researchers who were involved in gathering data would like to remain involved with the data after it is shared. The concern here is around a distance from the data, invoking a physical sense of being connected

and then disconnected with the set that has been shared – even though presumably they may have retained their own copy:

> *But, another thing which is, like I've said is, people don't normally feel like sharing data because ... it's partly because they feel they own the data, let me put it that way. That this is my data and I've worked so hard to get it so how do I share it in a responsible way that I still feel part of that data after I've shared it, so that it's not just out there.* ***I mean if I could get to continue to be associated with that data, not to be disconnected, once it is shared.***
>
> **Interview 5, Clinical Trials Manager, African, Research Institution, Africa**

This was then connected, by some interviewees, to an opinion that data or samples ought to remain in the place they had been collected, and indeed could be considered to be owned by the community or state – or, at the least, people from the place they had been collected ought to be involved in decisions about what happens to them. This was not only discussed in relation to the export of samples. The location that data were hosted was also raised by some from sub-Saharan Africa, as already seen in the section above, and again here in relation to Ebola data and samples from the outbreak in West Africa:

> *Suffice it to say, right from the outset of the data sharing, sample sharing, whatever activity,* ***I have been a strong advocate for the database, or whatever, to be hosted in Africa.*** *Ground servers in Africa, and to involve African management as far as possible. Arguments have been raised around the costs of this in Africa on account of costs of bandwidth, procurement and management of information, as well as personnel to carry out the job. Arguments have been made that it would be much cheaper to do it in [HIC university], given that some infrastructure already exists and there's some experience doing this kind of thing.* ***I still do maintain my position on my desire to have this done in Africa, because I think since this is where the samples came from, it would only be proper, and I think ethically right, to do this within Africa.***
>
> **Interview 28, Doctor and Scientific Director, African, Research Institution, Africa**

In the case of this interviewee, however, the reasoning was not about issues like academic recognition, but rather that as the burden of diseases and suffering had been in West Africa, there was a connection to the data and samples that meant they ought to be located there. He continued

to explain that the decision on where the data are hosted ought to account for factors beyond financing and economics, following a more non-consequentialist line of reasoning:

> *We don't make decisions on where to do things or carry out activities solely on economic grounds. We don't, you know, purchase the cheapest food. We don't wear the cheapest clothes. We don't make our decisions just based on how much it's going to cost us. There are several other things that go into, if I can put it in economic terms, choice models. So if we are looking at ... I mean, potentially, this database could be anywhere, including in space, because we have the ability to communicate with people in space. It could really be anywhere. The other question is,* ***what are the variables that inform the choice of location of this database? And those variables, for me, place a great premium on where did the samples that contribute to that database come from? Where was the greatest burden of suffering from the disease?*** *And all these other considerations I think are what need to be ... all these other factors or variables are what I think to be taken into consideration in the choice of where to locate the servers or where the data sharing should go on.*
>
> **Interview 28, Doctor and Scientific Director, African, Research Institution, Africa**

He also went on to say at a later point in the interview that he felt even more strongly that samples ought to be stored in Africa (as a continent, rather than the specific country they came from), linking it to "…this recognition of it being a part of a human being…". This relates to the point about samples being perceived as a part of somebody, as discussed in Chapter 4.

Some interviewees felt a dilemma resulting from this – considering that sharing widely and internationally could have a bigger impact, while at the same time having a desire to geographically restrict sharing, and share locally. The interviewee below highlighted this dilemma between global health impacts through sharing data being in tension with the individual researcher, with a feeling that the individual will lose out if they share over a wider geographical range:

> *Of course, sharing internationally has a much bigger impact, because then you have a wider scope. The default is that, naturally, people want things they can see. Then, there is that bias towards local sharing. Again, for me, it's a question of how you sell the whole concept so that someone doesn't see that, if it's shared internationally, they would probably lose much more than if it's shared locally.*
>
> **Interview 8, Clinical Trials Manager, African, Research Institution, Africa**

This is linked to the responsibilities to the wider population, and the perceived need to advance science on shared data and samples as rapidly as possible for others now and in the future who may be impacted by that infectious disease.

6.2.2.3. Seeking control to avoid negative assessment of data quality by others

Another reason for controlling data sharing, raised by a small number of interviewees, was that researchers were concerned about risks of exposing themselves to others with respect to their approach to data collection or analysis. One interviewee, involved in the development of a data sharing platform, suggested that some researchers may be worried that by sharing data, others will find their analysis approach to have been flawed:

> *The part of the fears is that when I do a clinical trial, I do the trials based on our internal piloting procedures, and in some cases people are not sure whether some of their procedures are up to their expected standards. So the moment to share the data out there, for example in the process of data cleaning, you do not clean data correctly there's a few errors ... a few discrepancies within the data.* ***If you share it out, somebody else may be able to pick out a few discrepancies from your own data, and it can make you look bad to see what you've published, it's published yes, but that data has a bit of issues.***
>
> **Interview 19, Data Manager, African, Research Institution, Africa**

This shows that while the opportunity for quality assurance and transparency of data analysis methods was given as one reason to share data, it may also act as a disincentive to share data.

While the quote above relates to the sharing of quantitative data, this is also shown as a concern when sharing qualitative data. The following interviewee described this in relation to her own research – her feelings of being vulnerable and exposed through the process of sharing data. She explained her view that those looking at the shared data may make judgements about her approach to interviewing without knowledge of the context in which she was interviewing, and the relationship she had established with participants:

> ***It does feel vulnerable to just put all that out there, or for people to look at the transcript*** *and think, "She was being a bit familiar," or whatever and not understand that we might have had a half hour conversation before, before the tape was turned on. Or not understand that this actually might be the fourth time that I've interviewed that person, or not understand that maybe I'm*

> *interviewing them in the park, or maybe I'm interviewing them in their home, or maybe I'm interviewing them in a cold clinical room and how that can impact the way you phrase questions, the way things are answered, there might be interruptions. None of that can be really properly understood by somebody else just coming in and looking at a transcript. It does. It feels exposing.* ***It feels exposing for me and it feels potentially exposing for the participants****. I think that that is ... That's a concern.*
>
> **Interview 27, Sociologist Researcher, Oceanian, Research Institution, Europe**

This also speaks to a concern about the distance or disconnect of those who may then go on to analyse the shared data from the data collection process.

6.3. Conclusions

This chapter has shown that interviewees considered themselves to have a number of responsibilities when getting involved in data and sample sharing. These responsibilities to different groups of people both promote and constrain sharing, with some interviewees feeling the personal weight of making a decision about whether data or samples ought to be shared, and what ongoing responsibilities they have in relation to the data or samples that have been shared.

Ownership and control mechanisms were discussed as ways to meet responsibilities, remain connected to the data or samples, and avoid negative judgements about the way data have been collected and analysed. Some interviewees based in sub-Saharan Africa also raised the issue of where data are physically stored, as well as samples, as being of concern.

An overarching finding of this chapter is that there is not a consensus around how to balance and meet different responsibilities, and interviewees with different backgrounds appeared to weight the responsibilities differently. For example, those involved in research or clinical care were, in general, more inclined towards protecting the people that data and samples were collected from.

The next chapter will consider a further theme arising from the interview data – how to enable ethical data and sample sharing, given the competing considerations and responsibilities presented here.

7. Enabling ethical data and sample sharing: The role and limits of consent, community engagement, trust and governance

The previous chapter showed that there are competing priorities with respect to data and sample sharing, and interviewees perceive themselves to have responsibilities to different groups of people. There is often a tension between these responsibilities, as interviewees generally considered it important to share rapidly to enable good quality research and facilitate the public health response, and at the same time felt a responsibility to protect the individuals from whom data and samples were collected. They were also concerned about loss of recognition and rewards from sharing. These appeared to be competing responsibilities working to promote and constrain sharing.

This chapter will focus on a related theme emerging from my analysis – the role and limits of consent, community engagement, trust and governance mechanisms to enable ethical sharing, from the point of data and sample collection, through to sharing and future uses. This theme builds on the exploration in chapter 6 about the competing responsibilities, and examines how in practice interviewees considered these could be resolved. I start this chapter by focusing on actions to be taken prior to sharing taking place (7.1). I then go on to examine perspectives on the role of trust (7.2) and governance mechanisms for sharing and future uses (7.3).

7.1. Actions prior to sharing

Consent, community engagement, and research ethics committee (REC) approval were discussed as actions that could or should be undertaken prior to data or sample sharing taking place. They are discussed in the following sub-sections.

7.1.1. Seeking consent

The focus of this section is on interviewee perspectives about the role and limits of consent for data and sample sharing in the infectious disease context, more specifically.

7.1.1.1. The importance of consent

Obtaining consent for data or sample sharing was often viewed as a part of meeting responsibilities to the people from whom data and samples had been collected. The previous chapter explored the responsibilities that interviewees considered themselves to have to the people from whom data and samples had been collected, and a number of interviewees felt that they had a duty of care for those people, and should respect their wishes about sharing. Consent was seen by many as representing the expressed permission of individuals that data and samples could be shared. This was often grounded in the traditional bioethical conceptualisation of consent, as in the following quote which related consent to the principle of respect for persons:

> *Yeah in ethical principles the corner…is to respect one's autonomy. To respect consent. That's why we are taking informed consent. We take informed consent because we respect the patient or participant.* ***See, if we are sharing the data without letting know about this process that is not respect for persons. You are not respecting the participant. So this will make it clearly unethical.*** *This, one of the basic principles of bioethics is going to be violated.*
>
> **Interview 9, Pharmacologist and Principal Investigator, African, Research Institution, Africa**

Where consent had not been obtained for ongoing storage, sharing and use, there was discussion about the need to destroy these data and samples. The following interviewee suggested that consent needed to be obtained for this at the point of collection of – in this case – samples:

> *Because you can't go back finding these guys because the rule of thumb is, get consent again. Where you didn't get it people are told, "No, destroy this sample. You didn't tell them you're going to do that. If it was not in the consent form, it was not in the blanket cover then it's too late." So most IRBs are saying and this is what I've observed and it's what I encourage, the PIs need to think through their projects very thoroughly at the beginning before they start collecting samples…*

Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa

It is notable that consent was not discussed as being required for the destruction of the samples.

7.1.1.2. Different forms of consent

While consent was considered important by a number of interviewees, the form of consent discussed differed, with some discussing specific consent for future uses of data and samples, and others discussing broad consent models. Broad consent models were seen as a way of reconciling the two competing responsibilities – to share and use data and samples as much as possible for knowledge and innovations that may be useful for public health, while at the same time ensuring that the people that data and samples were collected from have consented to sharing. It was also seen as a way of accounting for the difficulty predicting potential future uses of data and samples that may not be anticipated during collection:

> *I think we should move to a broader consent from the beginning which is we're taking the sample primarily for your health or, if we're not, primarily for this study, in which case of course it's very clear. It may be the case that it may be used for other health oriented research to benefit mankind. Not all of which can be predicted at this point.*
>
> **Interview 18, Doctor and Policy Maker, European, Government Funding Agency, Europe**

In contrast, a small number of interviewees felt that it was important that consent was given more specifically for other uses of data and samples:

> ***I think consent has to stretch all the way through. I think people have to give consent for exactly what happens.*** *You can't give consent for one thing and then something else happens. People have said, "I consent to my data being used for this purpose and this purpose only," then you have to respect that and not do something else with it that you never told them would get done. I'm not entirely certain why I feel that. It seems very important to me. I feel strongly about it.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

The same interviewee went on to say that consent was already generally "too broad" and that "you should resist any attempt to make it even broader." This seemed to be grounded in a concern that people would consent without truly meaning it – and this would not meet the aims of achieving valid consent. The interviewee did, however, also later discuss the limitation of obtaining consent for each use, and that it may be more practical to obtain a more general consent for putting the data in a repository. This shows how those involved in data or sample sharing, particularly as contributors, are torn between competing considerations – the need to share the data, but also the need to be clear about and respect the wishes of those from whom data or samples have been collected.

7.1.1.3. Consent – the importance of context

While interviewees generally saw consent as important, and gave a recognisable rationale for consent – that it respects the wishes of those that consent is being sought from - there were also apparent contextual factors that altered either the perceived importance or the practicalities of obtaining consent with respect to sharing data and samples. These contextual factors related to firstly, the emergency context; secondly, the different types of data and samples that may be shared; and thirdly, the cultural settings in which data or samples were collected. I discuss each of these in turn now.

Consent during emergencies

The acute response context was considered to add practical difficulties with respect to obtaining valid consent. While time pressures have already been noted in terms of the urgency to share data or samples quickly, they were also apparent at the level of individual interactions. The interviewee below discussed the limitations of obtaining consent from patients undergoing treatment during an outbreak – in this case in reference to Ebola. She described the time pressures for those working

in treatment centres who needed to maximise the time they were wearing personal protective equipment and their need to use this to care for patients:

> ***If I have people under PPE [personal protective equipment] and I try to have as little as possible because of the risks, should I put people only for the informed consent and putting them at risk only for the consent with all the things I've told you before about the constraints?*** *Or, should I have my caregivers spending 15 to 20 minutes explaining that because either you do a consent, or either you don't. If you do a consent it has to be done correctly. If it's bullshit, let's just drop it. If you want to have a real consent procedure where you explain everything, you need to think that it might take between 10 and 20 minutes. It's reasonable, maybe 20 minutes it's a lot, but at least 10 minutes, right?* ***You want them to spend 10 minutes in doing that, which means at least 10 minutes they will not be using to care for the patients.*** *You have to take that into account. We took all this into account, and we said no, we're not going to do that.*
>
> **Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe**

Obtaining consent for data and sample sharing was considered to be a lower priority task, noting that a full consent process is time consuming

The other issue was that for certain diseases such as Ebola, that cause a severe clinical picture, there was concern around the capacity of patients to engage with the consent process when acutely unwell, as explained by the following interviewee. She also suggested data sharing in not likely to be at the forefront of concerns for this group of patients, noting that in an outbreak they may have lost family members to the disease:

> *You know, when you have a fever of 39 degrees, and you're vomiting, and you have relentless diarrhoea. Your social support has ... half of them have died. Are you in a position to ask about your data? Not really.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

While the two issues above – time pressures, and acute illness – were noted to add practical difficulties obtaining consent, they did not mean that consent was considered unimportant. The following interviewee, for example, in fact considered it more important that consent be sought

for sharing samples (in this case) in an emergency setting. She was involved in the response to the Ebola outbreak in West Africa, working with a donor organisation. She related the importance of consent to the context – that people were very unwell and had little choice about presenting to the treatment centre and undergoing testing:

> *So I would, myself, view it as unethical for any of those samples to be used for anything other than an Ebola test. And also, bear in mind that walking into any of these centres or … well, the centres I suppose, was a horrible thing to do. A lot of people just didn't want to do it because the death rates in the centres were very, very high. People weren't sort of walking in just to have a blood test just for fun because it was interesting. It was a truly life or death situation and a very conscious decision to walk into one of these centres.* ***So that their understanding would have been, absolutely without a doubt, that their blood was being tested for Ebola. Therefore, to use that data sample for any other reason, I would suggest that you should get their consent to do it.***
>
> **Interview 3, Policy Maker, European, Government Funding Agency, Asia**

This shows a tension between the practical realities of obtaining consent and the ethical imperative to do so.

<u>Consent for different types of data and samples</u>

A further contextual factor that emerged as relevant to consent was the type of data or sample. Different types of data and samples were also seen as being more or less ethically permissible to share, as noted in Chapter 4. Some, such as the following interviewee, considered consent for sample sharing to be particularly important, as he considered samples to be more "complicated". This seems to be grounded in a concern about the potential for a wide range of uses with samples:

> *You know, when it comes to sample sharing that might make it a bit more complicated. So, when you take samples from participants regarding the informed consent … so, informed consent will have to mention "the sample which we are going to take from you is going to be used only for that purpose". And you have already, you know, you have already promised through informed consent process to the participant that the sample collected from this participant is not going to be used for any other purpose other than the one which you have consented for. Well, if you are sharing the sample with others, you don't know exactly what for other purposes it's going to be shared. So that might bring about even more serious problem.* ***Well, for me, data sharing might be better than sample sharing. Sample sharing could be very complicated.***

Interview 9, Pharmacologist and Principal Investigator, African, Research Institution, Africa

Distinctions were also made between the need for consent with routine and research data, with more importance given by most interviewees to consent for sharing research related data. The was discussed by the interviewee below, who felt that consent by individuals was not required for the sharing of anonymised data collected routinely through clinical interactions, but was required for that collected through research:

> *I suppose consent has to be central to any kind of research. Nobody can ...Okay, how do I say that? So if you're using public records, hospital records, you can analyse those anonymised. People don't need to give consent for that. The hospital, or somebody will give consent, but the individual won't on the basis that they gave that information particularly so that it would benefit them. I suppose what it is is…I'm saying* ***people do research because they want to benefit somebody else.*** *On the whole if someone's in a study, by-and-large, they may lose more than they gain and they're taking a risk. We're all taking a risk.* ***People give information in a clinical setting because they're going to benefit from it. They're…they're not doing it altruistically.*** *They not going to hospital altruistically. They're there because they need to go and whatever information they hand over is so that it will benefit themselves.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

This quote shows reasoning grounded in an individual risk-benefit calculation – that those taking part in research were taking on risks, without benefit to themselves, so ought to give consent before taking any additional risks to themselves through data sharing. Those whose data was collected through routine clinical care were perceived to have benefited from the interaction already, and so consent was perceived of a being less important to share that anonymised data.

There was not a general agreement between interviewees about the importance of consent for different types of data and samples, however. This interviewee, for example, considered obtaining consent for the sharing of research data to be potentially problematic. She was concerned that adding additional information on data sharing was distracting from the issues that potential research participants needed to focus on during the consent process for a clinical trial:

> *And I don't think adding a sentence, to be honest, in the informed consent about this data sharing will change much for those patients. At least in the countries where we are working, we try to keep our informed consent very simple, and we can see it's very hard to get a message through to those patients and so we want really to focus on what is really important for them in terms of ethics.* ***Adding more burden, more noise for something that they will probably not understand is not really an additional value.***
>
> **Interview 7, Clinical Trials Manager, European, Research Institution, Europe**

Other interviewees raised a similar concern, feeling research participants may consent to data or sample sharing without fully understanding what they are consenting to.

As noted in Chapter 4 a number of interviewees thought that consent ought not to be required for the sharing of aggregate data.

Consent in different settings

A further contextual factor discussed by a small number of interviewees was the appropriateness of obtaining consent at the individual level. While consent from individuals was the dominant model discussed, these interviewees (from different geographical locations themselves) discussed the diversity of settings from which data and samples would be shared during an infectious disease response. They considered the dominant model of consent to be grounded in 'Western' ethical norms not appropriate in other contexts. This is explained by the following interviewee, in reference to a country in sub-Saharan Africa. He suggested that in this setting consent wasn't conceptualised as something that only comes from the individual, but something that is obtained from community leaders:

> *So, consent has to be looked at within the perspective of the traditional system. What do I mean?* ***In our system, you cannot just go into a community to conduct clinical trials, or you cannot just go into a community to be able to collect samples. What that means is that there has to be layers of consent.*** *For example, there has to be administrative consent, and the administrative consent has to be obtained from the administrative structure. For example, the town chief, the community leadership, so you have to meet him, you have to discuss the processes that you want to implement, and you have to secure consent from them. That's part of our culture, that's part of our system. So that is an administrative consent. So you have to obtain that.*

Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa

This shows that while interviewees gave a general emphasis on the importance of consent, there were contextual factors – different factors discussed by different interviewees – that make obtaining individual-level consent challenging.

7.1.2. Community Engagement

7.1.2.1. The role of community engagement

Community engagement and/or permissions were identified by a small number of interviewees as a potential route to meet responsibilities to the people from whom the data and samples had been collected, through ensuring that people were generally informed about the potential for data or sample sharing. The term "community" was often unclearly defined but was generally seen as the people residing in the place that data and samples had originated from.

A small number of interviewees felt that the aims of community engagement went beyond informing, and that it was a way of seeking permission for sharing from the place that data and samples originated. The interviewee quoted here considered consent to be essential (section 7.1.1.3) but acknowledged the issues with obtaining it retrospectively – she was talking about using routinely collected samples from an outbreak. She considered community engagement to be an alternative to allow sample sharing and re-use to take place, given the challenges of obtaining consent during an emergency:

> ***I would expect quite a significant awareness campaign.*** *Based on all the awareness campaigns on Ebola that were made, certainly in [country], a small country, it is possible to have that kind of awareness campaign to say to people "this is the next step" and just to clarify, to make them aware. I think if people ... I mean, it was an entire population affected so if people felt strongly that those samples should not be used for further research, then I think you would know about it quite quickly and people would organise themselves to be able to make that case heard. And I would argue that if that started to happen, well certainly a discussion/negotiation should take place.* ***The will of any group that could potentially form should absolutely should be respected. I wouldn't personally want***

to see any kind of guarantee that further research should happen until that process had been undertaken.

Interview 3, Policy Maker, European, Government Funding Agency, Asia

The quote shows the interviewee saw engagement not as an instrumental tool, an opportunity to convince, but as a form of permission seeking – saying that whatever was collectively decided should be respected.

Community engagement was also seen by some as an opportunity to share knowledge that had been gained from using sharing data and samples. This form of engagement was seen as an ongoing and collective process, not something only taking place prior to data and sample sharing:

Ways to keep groups involved and in what happens to the samples, which they donated, this doesn't have to be on an individual basis either. One can think of how you could slightly alter this and expand it to a much bigger scale. And also, just make sure that they know even what's been learned, even if you kept it that simple, what's been learned.

Interview 18, Doctor and Policy Maker, European, Government Funding Agency, Europe

7.1.2.2. Concerns with community engagement

In a similar vein to the concern with the time it would take to obtain consent when treating patients while wearing personal protective equipment (section 7.1.1.3), the priority that should be given to discussions on data and samples sharing when engaging the community during an outbreak was also raised by the same interviewee (23). She highlighted the difficulties in engaging the community about data and sample sharing, noting that in a context of fear surrounding an outbreak it may be difficult to discuss this subject. She also went on to say that where engagement was happening with communities during an outbreak, she felt there were other priorities of things that needed to be discussed:

…we know that we made mistakes in health promotion activities in Ebola. We know that in the past, including in this one, it has had very violent reactions, including people dying, it's not a small thing. We

> *know that it's a lack of community understanding that made this particular outbreak spread up to Liberia was what we saw. We are understanding.* ***There is a reluctancy from the people who do operations to also include [data sharing] in the community discussions because of this fear.*** *And it is something that we should acknowledge. Now that doesn't mean we cannot work on how to prepare our packages and this and that. It might not be feasible at first hand, might be feasible only after a while, because it* ***[data sharing] might not be the thing that you want to start saying at your first encounter with the community.***
>
> **Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe**

A further concern that was raised was who could appropriately represent the community that data and samples had originated from, given the heterogeneous nature of members of a community. This interviewee, for example, felt it was not clear how a set of individuals could represent the plurality of people within a community – suggesting there is not one but many communities within a geographical space:

> *It's interesting this notion of community. In West Africa, or certainly in [West African country], if you were to say community, a community representative, you'd get a man. A lot of these data are from women, pregnant or lactating women. Women. Elderly people. That's not what you're going to get if you ask for a community representative****. I just don't know how well you advocate on behalf of other people when you are firmly not them.***
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

She went on to suggest that she would like guidance to clarify who ought to be there to advocate for a community:

> *And then you have a process whereby somebody who can advocate on behalf of these people whose data they are can be brought into the decision making. And again, that would have to be negotiated. Is it just a man? Is it a man and a woman? Is it a Muslim person and a Christian person? Is it somebody who's disabled? Somebody who's elderly? I don't know what that would look like, but somebody needs to make that decision. And were that in place, I'd feel a lot better.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

7.1.3. Research ethics committees

A governance mechanism highlighted by interviewees as potentially important prior to sharing was the process of REC review. As with consent and community engagement, interviewees generally conceived of this as a way of protecting the people from whom data or samples had been collected. It was believed that through seeking permission from RECs for the sharing of data or samples any potential issues and concerns would be assessed. None of the interviewees discussed RECs as having a primary role in enabling sharing to advance science, thereby meeting responsibilities to the people who may benefit from knowledge and innovations derived through sharing. The remit of RECs is generally such that discussions focused on either sharing data and samples that had been collected through research activities, or data and samples being shared for research purposes. RECs are not usually involved in sharing that is specifically for routine public health purposes, such as the surveillance of infectious diseases.

There was clear disagreement among interviewees about the appropriateness, and capacity, of RECs to be involved in decisions around sharing which will be discussed further below. Those based in low- and middle-income country (LMIC) settings generally saw them as essential to the process, while others, generally those based in high-income country (HIC) settings, were more sceptical about the value they added.

7.1.3.1. The role of research ethics committees

Giving permission for data and sample sharing

RECs were seen to have various roles with respect to data and sample sharing. The most commonly identified was their role as an authoritative body in making decisions to allow or stop data and sample sharing taking place. Some interviewees discussed them as protectors of the research participants and patients to whom the data or samples relate. The interviewee below outlined his view of RECs being critical in decision-making about sharing for this reason, going as

far as to claim that they were the primary protectors of the research participants, more so than the researchers themselves:

> *…the ethics committee are the ones who are concerned with protecting the rights of the people, although the investigator also should look for this protection but they are the ones who are responsible by law and by virtue. And they are the ones approving and overseeing and yeah, I think, it should critical part, yeah.*
>
> **Interview 11, Doctor and Principal Investigator, African, Research Institution, Africa**

Some interviewees discussed the permission from RECs as a form of consent, with their role being seen as particularly important when no consent for sharing had been obtained from the people from whom data and samples had been collected – for example for data and samples relating to studies that had already been completed. This is discussed in the following quote:

> *Well, I mean, consent, we have to…consent from the patient. Everything that we are going to do to them, and even their samples. Everything that is going to happen to them, everything in the trial, and even the samples that we draw from them. So we tell them about all that, yeah? I think it's quite unethical to do something that we didn't tell them in the consent form, but because usually the ... Maybe like, if they are trials that were done three years ago, it might be very hard to trace this patient. We cannot try and get them then to consent. I think this case,* ***these are cases that the ethics committee can look at and probably grant consent on behalf of the patient. Because they usually protect the patient.***
>
> **Interview 8, Clinical Trials Manager, African, Research Institution, Africa**

This interviewee, as with the previous interviewee, appears to be confident in the ability of RECs to protect research participants.

<u>Representing local views</u>

Another role of RECs that was discussed by some participants was to represent local views, recognising that data and samples in this context came from diverse geographical locations with distinct perspectives on the ethical issues. The following interviewee, a Laboratory Director based

in Europe, discussed this, suggesting that there may be different views of what is ethical. The implication here is that in-country ethics committees can provide these insights:

> *You could not do anything on the [West African] samples without their agreement, i.e. their ethical approval. And there might be different cultural views of ethics that you have to take into account because what's ethical to me might not be ethical to an African tribesman working in a pre-Christian, pre-Islam religious system. I wouldn't even understand.*
>
> **Interview 12, Doctor and Laboratory Director, European, Government Agency, Europe**

Develop ethics guidance

A further role discussed for RECs, by a small number of interviewees, was to develop a set of principles around sharing that could be used by those engaged in sharing:

> ***I think ethics committees should develop basic principles of data sharing.*** *And, well raise the ethical question which I might have some ideas but I'm not an ethicist so I think ethicist should be really flagging what are the ethical questions around data sharing? Then disseminate that within the ethical community and feed that back to groups like us doing data sharing and saying, "guys this is what you need to address".*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

7.1.3.2. Concerns about research ethics committees

Research ethics committees lacking capacity

As noted above, a number of interviewees, particularly those based in HICs, were sceptical about the ability of RECs to complete their roles – particularly where their role was to give permission for data or sample sharing. They expressed a lack of confidence in the REC system. For some interviewees this was in relation to specific RECs who were not seen as having the capacity, that is, the knowledge and skills and/or time to adequately make judgements on the issues around data and sample sharing. For others it was considered a more general issue with RECs not being well versed as sharing was seen as a relatively new approach. The interviewee below discussed obtaining

ethical approval for research during the Ebola crisis – and suggested a lack of confidence that having ethical approval was an indicator that the research was being conducted in an ethical way:

> *…I'm not saying that this expedient ethics approval process during Ebola was necessarily wrong, because it's certainly the case that they needed ... They were under a lot of pressure, and they needed to do something. They couldn't just say, "Well, we lack the capacity, so we're not going to do anything." And* ***I think they were probably trying to do their best, but insofar as that actually is an ethical vetting of these applications, I'm probably pretty sure it's not what we would expect.*** *Yet it probably appears like it is, and I think people treat it like it was. Like, "Oh, look. I have ethics approval. Everything's fine." When in fact, that's not what it means.*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

The interviewee highlighted a tension between achieving best practice with respect to in-depth ethics scrutiny, and allowing research to happen quickly in the face of an outbreak.

One issue for RECs based in more resource constrained settings, raised by the following interviewee based in sub-Saharan Africa, was having the appropriate skill-set represented on the committee. He felt there were a lack of people with a background in bioethics, and those few that had had often emigrated, leaving RECs based in these settings facing difficult questions, but with a lack of people with the training to discuss the issues raised by data and sample sharing:

> *…in [LMIC], I think in terms of ethicists, one who has, say, a masters in bioethics, PhD, bioethics, we do not have any [West African] in country that have a PhD in bioethics. There are a couple [West Africans] that are doing masters in bioethics, none have returned home.* ***So what you have, you have an institutional board with [West Africans] that have been involved in numerous short - one week, two week, three week courses in ethics, so how do you compare those ethical board with the ethic board at [HIC] University. There is a big disparity.*** *So what that means is that these limited resource ethical boards may not be in a position to better discuss issues of intellectual property, better discuss issue of royalty, better discuss issues of consent, relative to the one at [HIC] University where you have so many people who are highly knowledgeable and the system is well structured.*
>
> **Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa**

The quote shows a further dimension of potential inequity between HICs and LMICs, when sharing data and samples.

Research ethics committees slowing things down

A further concern with RECs was that they would slow down the process of data and sample sharing, and this was considered in tension with the public health aim of rapid, early sharing. The next quote illustrates this, outlining the implication for this interviewee that patients from whom samples had been obtained would then not benefit from results obtained from research on shared samples:

> *But I suddenly think that it's sometimes a bit overprotective and also not in the benefit of the patient sometimes. For example, if a data analysis could be done quickly without re-applying for ethical approval in my opinion, sometimes unnecessarily just because the regulations says so which takes sometimes six months and maybe the patient cannot benefit anymore from that analysis.* ***If we had been able to do it quickly, maybe we could have still, like, implemented a new treatment for the patient or have different perspective for the patient's future.*** *So I think it can also, yeah, have a down side to having overprotective regulations.*
>
> **Interview 4, Pharmacologist Researcher, European, Research Institution, Europe**

This is in the context of a disease such as Visceral Leishmaniasis, with a more protracted clinical course in which the patient may benefit from new research, rather than a short-lived in infection such as influenza. The public health aim to share rapidly and early also included sharing that would benefit other people (as opposed to those that the data and samples were collected from).

Too many research ethics committees involved

A further concern raised with the general system of RECs approval was that sometimes multiple approvals were required. Different RECs may come to different conclusions or give different advice, leading to difficulty moving forwards to share the data and samples:

> *What happened is a lot of researchers or whenever you get clearance from the [research ethics committee] and your project requires samples to be transferred across the counties within the country or across nations, a lot of researchers get frustrated after getting IRB approval, they have to come to [East Africa] and get approval at different, random, often sometimes not clearly cut out.* ***The pathway to getting all the clearance needed until your sample gets to where you're going is not very clear.***

Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa

As a result of the aforementioned concerns some interviewees envisaged a more limited role for RECs. That of being kept informed about data sharing activities related to research, rather than giving permission for this. This interviewee below discussed this – she did not consider RECs to be responsible for approving individual uses of shared data. She goes on to explain her concern that by asking for permission, rather than merely informing, RECs are in a position to halt data sharing – although it is promoted by bodies such as journals:

> *I think we should not resubmit each [data sharing] project to the ethics committee separately. It should be something much more broad to explain. I think it goes together with a request at the international levels, from the journals. So I don't think that they could refuse this. I would have been in favour of informing them more than asking for an approval, because they are always able to say no if something is really not fulfilling their request.*
>
> **Interview 7, Clinical Trials Manager, European, Research Institution, Europe**

This quote highlights a further consideration – reconciling ethics approvals with the requirements of funders and journals to share data.

7.2. Trust and confidence between professionals at the point of sharing

The previous section focused on mechanisms and actions taken that would allow, or not allow, data and sample sharing to take place. Once sharing and onwards use were taking place, interviewees primarily discussed the role of trust between sharing professionals, and the role of certain governance mechanisms, as a way of undertaking sharing while accounting for identified ethical issues. There is a large body of literature relating to trust, and different conceptualisations of trust. The findings presented here are grounded in the way in which interviewees discussed trust

and confidence in sharing – I have not used a specific definition of trust with respect to these findings.

7.2.1. The importance of trust

Trust between those engaged professionally in sharing was seen as an important, if not essential, requirement for data and sample sharing by a number of interviewees from a range of backgrounds. Trust was variously described by different interviewees and was usually (but not exclusively) discussed at the interpersonal level. Trust was seen as a way of enabling data and sample sharing, as there was a sense that the data and samples would be looked after when sharing with known, trusted others:

> *Well your friend will lend you a jumper because he is your friend and he trusts you that you will take care of his jumper. So it's a bit the same situation that you want to share the data with somebody that you believe will take care of the data and will not misuse and will acknowledge that actually can do whatever he does thanks to you in the first place having in the first place facilitated data sharing. Or you get warm because your friend has loaned you the jumper.*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

The interviewee discussed concerns that people have around data and sample sharing – that the data or samples will be misused, or that they will not be acknowledged for their work – and described trusting relationships as a space where these concerns, which have ethical implications, are not there.

A further issue of importance for interviewees was ownership of the data and samples, as discussed in Chapter 6. Trusting relationships were also seen as a potential way for interviewees to share and transfer custodianship, and cede control of the data and samples:

> *I think, I would say that it's a bit complex to say who's explicitly the owner of the data. So that before the data is shared, of course, the person who has collected that data. Either the investigator or the sponsor is the owner of the data set. Then when they share it, I think they still have some sense of ownership on it, but it's probably not full ownership, because it's already out of their hands. Then, their ownership to it is actually out of trust that, whatever agreements you have with them will be kept.* ***So that even though***

> ***you become the custodian of that data, they trust you that you own it on their behalf. It's like transferring that responsibility.***
>
> **Interview 13, Statistician and Group Head, African, Research Institution, Africa**

Some went on to describe trust as an essential requirement for enabling data and sample sharing to take place. It was also seen by most interviewees as something that required time to build up. This interviewee discussed his experience developing a data sharing platform, with those contributing data moving from a reticence to share data to a point at which they were happy for data sharing to take place without their involvement. He attributed this change to the development of trust between those contributing their data and those managing the data repository:

> ***We have built trust with them and this perception of trust is essential in the data sharing process.*** *I think or research groups which have been contributing to their data, some have become ...* ***From being fairly reluctant at the beginning they are the one now asking to share openly their data.*** *So we have seen an evolution in some groups where they essentially say, "oh don't bother me with your request of sharing my data, you can share it with whoever you want, I don't care. Sort of…I see the value and I know that I would not do that much with this data anymore".*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

Trust was linked by some to a sense of familiarity with the other people engaged in sharing, by the interviewee quoted above and others. Conversely a lack of trust was seen as a reason that people may not share data and samples:

> *Yes [trust is] a very critical component of the data sharing process. Without that then, either you end up with a situation where you have people not willing to share data. Or, even if they are willing to share, they take forever to share. Which again, beats the whole point of having pooled data as early as possible.*
>
> **Interview 13, Statistician and Group Head, African, Research Institution, Africa**

As noted in Chapter 4 some forms of data and sample sharing may involve sharing with people who are not known to the contributors, for example, when sharing through a repository.

7.2.2. Trust and location

A small number of interviewees linked trust with the location of the people they were sharing data or samples with. For example, this interviewee considered trust to diminish as sharing took place with those further away. One interviewee linked this with a sense that people felt they were losing control and felt threatened – sharing with nearby countries was seen as being less threatening and this was equated with "trust" sharing there:

> *I don't know how to put it in a very nice way, but I think it boils down to the issues of trust. I mean, like, I think … Let's say the people in Uganda ... For example, we handle the Uganda data, Ethiopia, and even Sudan connection. We handle it here. I think it's majorly because of trust. I think people don't feel threatened, yeah.* ***When it begins going farther and farther away from them, and many other people are getting involved, I think the issues of trust emerge, and also people feeling threatened at the loss of control.***
>
> **Interview 8, Clinical Trials Manager, African, Research Institution, Africa**

She went on to discuss her particular concern that those in HICs would be able to rapidly analyse the data, and those contributing the data may lose involvement. The nearby countries this interviewee discussed are also all part of a well-established network researching an infectious disease, and so the relationships are well developed, and this may be related to the interviewee's perception of these two different sharing situations.

The following interviewee also directly linked geographical location to willingness to share, and discussed trust in relation to countries rather than individuals or groups of researchers. While he used the word "trust", this seemed to be about a lack of confidence in developing a good collaboration, given the political hostility between certain countries. He discussed his own experience that samples would not be requested from countries where there may be hostilities:

> *I mean there has to be a sense of trust between different countries, and I think inevitably there's sort of international suspicions and so on, and those are not the countries that are likely to collaborate. And, you know, I think right from the get go, you know* ***we wouldn't look for specimen in countries where we thought they would be hostile to these kind of things.*** *So, I'm sure there's a lot of trust that's involved.*

Interview 25, Doctor and Government Worker, African, Regulatory Agency, North America

Both of the above examples, which give a sense that there are preferred geographical locations for sharing data or samples, can be seen as a clear challenge where sharing may need to take place across a number of geographical locations during a global infectious disease emergency.

7.2.3. Gaps left by trust

While trust was seen by a number of interviewees to be an important factor in enabling data and sample sharing and mitigating concerns, it was also considered problematic in various ways by some. Trust was seen to be challenging in emergency contexts – as noted above trust was thought to be something that takes time to build, and where groups of people and organisations come together without a pre-existing relationship, as in certain outbreak responses, trust is not likely to exist between these people. One interviewee considered it still possible to share during such emergencies, and she described this as "trust on the hoof":

> *I think, through that,* ***you can build up trust on the hoof****, because I think in that situation everyone's in the crisis together. So it's not artificial, but it's slightly different than the norm. You're all in this crisis together, you're all keen to do what needs to be done as long as it's done correctly to make things better.*
>
> **Interview 22, Doctor and Group Head, European, Research Institution, Europe**

The sense here is that the goal orientated response to outbreaks makes people come together and share data and/or samples, as if in a trusting relationship.

A further issue suggested by one interviewee was that trusting relationships would not necessarily develop, even over time. She recounted her experience working on a project involving the sharing of dried blood spots, and that a team in a neighbouring country, who had worked with the other team for a period of time and appeared to have a good relationship with them were nevertheless not willing to share the dried blood spots they had collected:

> *These are dried blood spots, so sources of DNA from parasites. They did that and it was decided it was going to be in [country 1] okay? The [country 2] never sent a sample even though the idea was you had this filter paper and you'd put three spots on it. One spot went to [country 1] and two spots stayed in [country 2]. Never happened.* ***I thought, "Wow, these guys that are working together for a long time, they all know each other, they really do trust each other pretty much." No way.*** *I think those, I don't know why but it's different and people feel much more tied to those samples from [country 2] or wherever they're from.*
>
> **Interview 16, Scientific Director, North American, Research Institution, North America**

The quote also relates to the discussion in Chapter 6, and the desire of data or sample contributors to remain linked to them. This desire to be linked looks, in the quote above, to be overriding any trust that may have built up that would enable sharing of the samples.

Trust was also seen as potentially unreliable, as discussed in the quote below. The interviewee felt that relying on trusted relationships with known others risked people taking less care to stick to the rules. She discussed this in the context of someone sharing identifiable information with her that she felt was inappropriate, and considered rules to be important to preventing this:

> *I think, even when we're sharing data within little groups like that, it's…you think everyone trusts each other. We're all nice people. That's not a good enough reason to do these sorts of things. You can't say, "It's okay because we're all nice people." Because* ***that just opens the door for mistakes to happen and you've got to have rules and people have got to stick to them****.*
>
> **Interview 2, Epidemiologist Researcher, European, Research Institution, Europe**

7.3. Governance for sharing and future uses

While trust was seen as important or even essential for effective sharing by a number of interviewees, some interviewees also felt that using trust as a model for sharing was potentially problematic, as noted above. This section focuses on the theme of governance of sharing and future uses of data and samples. Two specific governance processes – agreements, and access committees – were most often discussed by interviewees, and are explored in more detail.

7.3.1. The role of governance for sharing and future uses

Governance mechanisms were discussed as a way of enabling data and sample sharing, in a way that accounts for identified concerns, and seeks to meet responsibilities to different groups. Such processes were seen as ways of delineating roles, rights and responsibilities when sharing. This is outlined by this interviewee, who considered such governance mechanisms as a way of enabling sharing while being accountable to others:

> *I think the collaboration really is to explore all those areas, especially if the areas are challenging in terms of ... its dynamic, it's a field that is changing and be able to foresee some of those areas so that they actually put protections for that as much as possible. Of course, not refusing to share the data but, just making sure that there's a good framework, whether there's a memorandum of understanding, maybe a contractual agreement, whichever level it is, they do it. And* ***everybody knowing the responsibilities and who is, for example, like you said, accountable in case of a breach.***
>
> **Interview 5, Clinical Trials Manager, African, Research Institution, Africa**

Some form of process or governance mechanism was also seen to have a role in managing sharing between HICs and LMICs, relating to concerns about differences in power, as noted in Chapter 5. This is discussed by this interviewee based in Africa, who discussed ethics frameworks as a tool to use to prevent "abuse" of people, relating this to previous abuse of those based in LMICs by others coming in to do research in the setting:

> *[LMIC] historically has been a least developed country, historically abused. Ethical frameworks are coming into place because we know numerous instances where population has been abused so many times. I mean, that's what you say, I don't want to name some of the instances in historical context, but they happen so*

many times. So I think mechanism has been put in place to be able to try to minimise these abuse, minimise these disadvantages…

Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa

The use of the word "abuse" gives a strong sense of historical injustice, and its relevance to current data and sample sharing practices for this interviewee.

A general concern with governance mechanisms being implemented when sharing was that it would ultimately limit sharing. A number of interviewees considered data and sample sharing as important to develop knowledge and innovations, and there was concern that this would be impeded by complex rules and governance mechanisms:

Now, you need to be careful that it's not going to be a limiting factor for people collecting data and sharing it. It's always a sort of balance whereby we do put rules, but the rules should not prevent you then at the end, because it's so complex, that at the end nobody does anything anymore.

Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe

The two types of governance mechanisms there were discussed when engaged in onward sharing and use of data and sample – agreements and access committees - are explored in the following sub-sections.

7.3.2. Agreements between sharing parties

7.3.2.1. The role of agreements

The use of agreements such as data sharing agreements (DSA) or material transfer agreements (MTA) between sharing parties was discussed by a number of interviewees. Agreements were seen as a potential way of meeting responsibilities to the stakeholders in data and samples sharing – other professionals, the people from whom data and samples had been collected, and the wider population who may benefit from data and sample sharing. They were, however, most often

discussed as a way of protecting the professionals who were data and sample contributors. They were considered as opportunities to lay out and agree upon the expectations of the sharing parties, and approaches to issues such as intellectual property rights and the distribution of benefits arising from research. The role of agreements – in this case MTAs – as a way of ensuring the sharing of benefits between data contributors and users is discussed in this quote:

> *That why you need to have material transfer agreement. The material transfer agreement, it has to be clearly explicit to all these things, you know. If, for example what you really want is one scenario…If there is any product coming out of this…then the benefit which is obtained from that product should be shared between the donor and the receiver. That could be one good example, you know. Which you can get from the material transfer agreement.*
>
> **Interview 9, Pharmacologist and Principal Investigator, African, Research Institution, Africa**

Agreements as a way of enabling control and providing boundaries

As discussed in Chapter 6, the loss of control of data and samples was a great concern to a number of interviewees – and this was a potential reason that interviewees may be reticent to share data or samples. Agreements here, as illustrated by the following quote, were seen to have a role in alleviating some of the fears from data contributors that they are losing control of the data:

> *Yeah, you can have documentation. People have agreed to having the data can be used for this and this…People enter into agreements that should be able to reduce those kind of ... Because I think, in my opinion, there's a bit of suspicion. Yes,* ***there's a bit of suspicion. People fearing to lose some form of control. If people enter into some clearly written agreements, maybe that could help.*** *If people stick to agreements, something like that.*
>
> **Interview 8, Clinical Trials Manager, African, Research Institution, Africa**

The quote shows that agreements are seen to provide boundaries for use of the data, thereby enabling some control on the part of the data contributors.

Agreements, intellectual property and the distribution of benefits

A particular area of concern for interviewees who were data or sample contributors was the distribution of benefits from research. Intellectual Property (IP) claims on innovations from shared samples in particular were seen as a potential threat. The concern was that they may professionally lose out as royalties are then claimed by the sample users, and also concern that this would impact on people who may benefit from innovations being able to access them. The interviewee below discussed agreements as a tool to prevent IP rights then being applied for without prior agreement from the data platform. In this case the agreement between those managing a data repository or biobank and its users, rather than the initial data contributors (unless they are represented) or the people from whom the data or samples were collected:

> *In the context of Ebola, because of the partner that we are working with there's a lot of fear about intellectual property rights that could be abused after the fact. We are trying to have a language where we say to a third party as an obligation which will be signed on a material or data transfer agreement. But if intellectual property wanted to be licenced on the basis of this work then they would have to come back to us for negotiation.*
>
> **Interview 6, Doctor and Research Director, European, Research Institution, Europe**

Concern around distribution of benefits, and the use of agreements to manage this, was also discussed at the national level by a small number of interviewees. This interviewee discussed the development of an MTA by a National Bioethics Committee in East Africa – not specifically for infectious diseases but for use across a range of areas that would include infectious diseases:

> *Looking at what the National Bioethics Committee was doing by developing the material transfer document because it's connected in other legislation around transfer of bodies and samples from the country. I'm not very sure but I know the reason they did that is just to protect indigenous knowledge but also, you know, indigenous samples that may be unique to the country. And just to be sure when anything goes out of the country, it would not come back to bite someone's butt really. That's basically it.* ***We're never really sure what the problem can be but also there's the idea of the benefits that may come up may be so huge that the country wants to benefit from that or to be recognised for it.***
>
> **Interview 10, Research Ethics Committee Member and Social Scientist, African, Research Institution, Africa**

The main use of the MTA described here is to protect knowledge and samples originating in the country in question, and ensure the country benefits from sharing the samples.

7.3.2.2. Concerns with agreements

Emergencies as a challenges to the development of agreements

Emergencies were discussed by interviewees, predominantly those who were involved in developing data sharing platforms, as a particular challenge to developing agreements – because prior agreements may not be in place, and the situation could be chaotic. The need to develop agreements in advance of an epidemic is discussed in the following quote:

> ***In a crisis like that in an emerging infectious disease of any kind unless you have some procedures in place before the fact, I don't know how you handle that.*** *Yeah, I don't have any really good ideas. Once it's started to happen and it unfolds in the tragic way that particular epidemic unfolded, yeah. That's why, you know, [anonymised] and lots of other people are trying to get things, sort of, procedures and agreements, in place before the fact in a generic way so that whether it's an Ebola or SARS or whatever, whatever disease that suddenly erupts, you'll have a structure in place for all of these things in principle.* ***The problem is you don't know where it's going to happen*** *so you could work with the governments of, you know Liberia, Guinea, and Sierra Leone and have it all set up and it never happens there again so it's not that useful.*
>
> **Interview 16, Scientific Director, North American, Research Institution, North America**

The interviewee noted that developing agreements in advance is difficult as it was not clear where the next emergency will take place.

MTAs are also legal documents, allowing the transfer of samples between jurisdictions, according to the stipulations set out in the agreement. Some interviewees were concerned that in outbreak situations, the chaos of the outbreak meant that some groups may not engage with developing agreements, particularly where procedures are not in place in advance. One interviewee, an African public health doctor, gave an example of sample exportation occurring without agreements in place during an outbreak:

> *…there was a proliferation of organisations, there was a proliferation of institutions that came to help the country. But because there were no guidelines, no mechanism that was in place at the time, no structure, no governing structure that was in place at the time, so you expect that. The way the emergency was there was chaos. The way emergency during crisis situation situations, you expect that.* ***So all I can say is some international institution took advantage of the chaos, took advantage of the emergency health situations, and took some of these samples out of the country.*** *So those things happened during emergencies.*
>
> **Interview 21, Public Health Doctor and Researcher, African, Government Agency, Africa**

Agreements not adding value

As with other governance mechanisms, agreements were considered, by some interviewees, to have certain limitations – with the nature of the limitation grounded in the nature of the agreement and the context it was being used in. The first was a concern that agreements did not add value. This concern was expressed by the following interviewee, who felt that much of what was noted in the MTA, issues such as maintaining confidentiality of research participants, was often covered in other contracts held within the collaboration he was working in:

> *To be honest, I never cared so much about [material transfer agreements] because most of the statements in them, for example, regarding confidentiality or regarding the transfer itself, that is already maintained in the separate contracts that we have for example with [the initiative] or with the respect of universities that we are collaborating with so yeah for me,* ***the material transfer agreement seems a little bit double, extra agreement on top of another agreement. For me, personally, it's not really useful but I guess from a legal perspective it's probably required.***
>
> **Interview 4, Pharmacologist Researcher, European, Research Institution, Europe**

The relevance of this concern is contextual and will depend on the nature of the MTA and other contracts that are held between those engaged in sharing.

Agreements adding burden to the process

Some interviewees raised practical issues with certain agreements that are tailored to recognise ownership and control of data contributors, by making them the permission-givers for onward

sharing. The interviewee below, who has extensive experience developing a data sharing platform, highlighted this, suggested that while such agreements allow the data contributor to retain control, it was also burdensome, as the contributors had to be contacted for each use of data, and at times they were unable to contact them:

> *The agreement at its base is that those data belong to the people who collected them and that any use of the data must be cleared with them before it's included in one of these proof analyses or whatever. And that has worked well but it's of course not open data sharing and so we're in process now ... Because* ***that's turned out to be logistically an incredible pain, you have to ask every single time and many of these data sets have been used to ask a lot of different questions.*** *And so each time the people here at [HIC-based university] are doing these various different scientific administration things, chasing studies as they say, have to go back and they have to get an answer. At some point the email is no longer operative or whatever so there are lots of practical things that make that difficult.*
>
> **Interview 16, Scientific Director, North American, Research Institution, North America**

This again raises the question of who ought to control data and sample sharing. Agreements here become an expression of the verdict on this question, and the decision taken has practical implications for how data and sample sharing is conducted. There is the potential in this example of an ethical issue to arise if worthwhile work on shared data and samples is delayed as a result of this agreement, with implications for meeting responsibilities to the wider population who may benefit from work on shared data and samples.

From the perspective of a number of researchers in this study, however, many of whom were relatively new to data sharing, their involvement in decisions around the sharing of data or samples they had collected was desirable. This interviewee discussed her own preference that she be contacted if people wanted to use the data, and invite her to be involved in its use, seeing it as an expression of the recognition of her role as a data contributor:

> ***I think I would be happier if there was requirement of people interested that they had to contact me in the first interests and I had to be involved.*** *If someone said to me, "I really interested in the data, I'm really interested in writing it up, interpreting a different way or coupling it with other data from a comparable study, is this okay, would you like to be involved," and then for me to be kept abreast of what's happening and then be involved in authorship, I'd be all right with that. And*

that would be nice. It would be nice to be able to collaborate with other people, get new perspectives. Also to have a sense to be able to say, "I'm concerned about this." And then the recognition. To be included as an author. And that's not only about contributing to the paper but it's a recognition that you devise the original study and collected the data. So I would be okay with that.

Interview 27, Sociologist Researcher, Oceanian, Research Institution, Europe

Agreements with known and unknown people

While agreements were seen as a tool to manage the concerns of data and sample contributors, in the case of this interviewee he did not have confidence that an agreement would be developed for more widespread sharing internationally, that he was being asked to participate in:

Yeah, because this one we are working with people that don't know all of them, that's where complications come from here. We don't want ... The one we are working with now, I know them all. But for the international one, so the [HIC] hospital, no, I don't know. That's not alright. With the one I'm working with now, we have an agreement and everything. Nobody can do something unless we all sit together and agree on it, but for the international, that will not happen.

Interview 11, Doctor and Principal Investigator, African, Research Institution, Africa

This quote again underscores the importance of relationships for data and sample contributors.

7.3.3. Access Committees

7.3.3.1. The importance of access committees

A further governance mechanism discussed at the point of sharing between professionals was the use of access committees to make decisions on who can access the data and samples, and for what purpose. As discussed in Chapter 4, the interviewees in this study generally did not advocate for fully open access data and sample sharing models – with managed access being seen as a way of protecting the data and samples and guarding against misuse. This is discussed in the following quote:

…we also thought that as part of policies on sharing data, we should have data access committee, an independent body that will be able to look at the different requests to the data that has been shared in the

repository, and be able to grant access to such requests based on, you know, solid evidence or other proof from the requesters on why they need the data and what they want to do with the data, so that we don't have a data sharing platform where people can just dump in their data and once the data is dumped in the data sharing platform, then anybody else can be able to get hold of this data and do anything with this data in an uncoordinated manner. These are some of the things that at the end of the day will make investigators uncomfortable because they will not know who is doing what with the kind of data that they've being shared. But ***if we have a coordinated approach where we have a data access committee, reviewing all the data requests and granting access to such requests, then it makes it a more responsible data sharing platform, where I will only go to ask for data if I have, you know, a proper reason for asking for that data and for using that data.***

Interview 19, Data Manager, African, Research Institution, Africa

The interviewee described data access committees as a mechanism to make data sharing "responsible" – that it is coordinated, and data are used for "proper" reasons. This enables some form of control to be exerted on the data sharing process

None of the interviewees questioned the utility of access committees, although a minority saw them as a tool to manage the concerns of data and sample contributors rather than being intrinsically valuable.

Access Committees enabling control

As with agreements, access committees were suggested as a way of controlling data and sample sharing (they were considered necessary for samples) and allaying the fears of contributors about shared data and samples being used for bad research. The following interviewee discussed the use of access committees (she termed them scientific advisory committees) as a way to acknowledge and respond to concerns of data contributors that data will be misused. Unlike other interviewees, however, she suggested that this was an interim step, as she considered it was not in keeping with the ultimate goal of widely accessible data:

…I think what they have proposed is to have still some kind of control on who would access through a scientific advisory committee. They were talking about other subcommittees that would be in charge of … for viewing the proposals and allowing or not the access. I think it's probably not the goal at the end because probably at the end we would like to have most of people accessing to the data, and … But ***it's a first***

step which probably is necessary to reassure the researchers that we are not going to do something completely wrong with their data.

Interview 7, Clinical Trials Manager, European, Research Institution, Europe

Access Committees to represent views of different groups

Access Committees were also seen, by a small number of interviewees, as an opportunity to represent the views of different groups with respect to onward sharing. The interviewee below discussed ensuring that the individuals from whom data and samples were collected are represented when it comes to decision-making with respect to data sharing:

These should be people who are not necessarily just involved with that particular outbreak if that's why the samples ... They should be people who can represent that individual, and so there should be people who can interpret new data they come along, understand the implications for health, so can make a judgement and of course, you know, representatives ... People who can represent that patient or whoever, that participants view, so in general that's lay members of a committee, across the board really. In fact, probably less important to have somebody who's involved specifically with that disease for which the samples were received, because that's much easier, because consent would have been obtained in general to study that disease. But it's the other aspects that are probably more important.

Interview 15, Doctor and Scientific Director, European, Research Institution, Africa

This may raise a tension of using data and samples with a view to what is optimal or best for global public health, versus using them in accordance with the wishes of those representing the individuals from whom they were collected, who may have different views.

Some interviewees considered independence to be important with respect to the committee:

So, in emerging infections our opinion is, this is a general rule for data sharing, this is independent from emerging infections. When it comes to epidemics we are believing that another thing that could help is to work with independent data sharing platforms that have these kind of rules. ***Then the ideal situation would be to have really, I mean for as much as independence is something reachable, so independent, neutral, no conflict of interest body that sort of manage the data sharing platform, that decides on the data access terms and conditions and then follows it up.*** *In the super-best case scenario, could pin down whoever is not following the rules. Of course once you give the data, there needs to be a sort of follow up where people are compliant.*

Interview 23, Doctor and Adviser, European, Non-Governmental Organisation, Europe

In this case independence was discussed as neutral and lacking conflict of interests, which may restrict the involvement, for example, of data and sample contributors who may be reticent to share with competitors.

7.3.3.2. Concerns with access committees

Beyond considering the composition of the access committee, there were few concerns raised about their role, and those that discussed them generally considered them a useful way of ensuring data and samples are shared for good research by people who were appropriately qualified. One interviewee who was a data contributor raised a concern, however, that she felt a lack of clarity around the delineation of responsibilities (as noted in Chapter 6), feeling unsure about whether she had ongoing responsibilities for how the data were being used, although the decisions about this were being made by committee:

> *But it concerns me a great deal, this issue of data pooling, or data sharing with a central repository. Governance structure there, and at what point does our responsibility end?*
>
> **Interview 26, Researcher and Adviser, North American, Non-Governmental Organisation, Europe**

7.4. Conclusions

This chapter has shown that consent, community engagement, trust and governance mechanisms were often perceived to be potential solutions to enable ethical data and sample sharing, that is, sharing that accounts for identified responsibilities to others. In particular, they were seen as ways to allay the following concerns when participating in data and samples sharing:

- That it would be done without respecting the wishes of, and adequately protecting, the individuals and communities from which they were collected;
- That contributors would not be acknowledged for the work required to collect the data and samples;
- That it would lead to poor quality research, or research that is denigrating or stigmatising to particular groups of people;
- That the distribution of benefits from knowledge, or innovations deriving from the shared data and samples, would not be fairly distributed.

All of these potential solutions were also seen as having differing limitations in achieving the above aims, and so could be seen as working in tandem. For example, consent was noted to be particularly difficult to obtain during emergencies, and community engagement and RECs were seen as other routes to enable sharing to take place in a way that was perceived to be ethical, despite a lack of consent. Similarly, trust was seen as on the one hand essential by some, but by others as potentially restrictive or a risky way of enabling sharing, and governance mechanisms such as agreements and access committees were then seen as a way of formalising expectations and protecting against misuse of data or samples.

The emergency context was perceived to add additional challenges, such as the time pressures both at the individual and system level, making it difficult to obtain consent, or feel confident that a robust ethics review had taken place by a REC. There were also other priorities discussed as relevant to an emergency that meant that engaging people about data and sample sharing, at the individual or community level, may be perceived to be a lower priority. Data or samples may also need to be shared across a wide geographical area, with some interviewees expressing concern about which countries are and are not involved in the sharing.

A further notable finding in this chapter, that is shared across Chapters 4 to 7, is the wide range of perspectives and lack of consensus among interviewees. Some felt that individual level consent was vital, whilst others considered that this was not an appropriate model for some settings. Some felt community engagement was being done to seek permission, whilst others felt that it was merely a way of informing communities that sharing was taking place. Some felt confidence that RECs were the primary protectors of research participants, whilst others felt that they were not necessarily well informed about the issues at stake. Some felt that trusting relationships were essential, whilst others felt that this was problematic basis for sharing, and it was better to operate with clearly specified and agreed rules.

The next chapter brings together the findings across Chapters 4 to 7, for further interpretation, and examination in relation to the literature.

8. Discussion

8.1. Introduction

Infectious diseases pose an important global threat, as has been made clear following outbreaks of SARS, influenza, Ebola, Zika and now, at the time of writing, COVID-19. As noted in Chapter 1, infectious diseases are a broad group, including endemic as well as epidemic-prone infectious, which are transmitted through varying routes. Given the potential for some infectious diseases to spread rapidly and widely, international collaborative action is required to respond to infectious disease threats. Data and sample sharing[17] are discussed in the literature as ways to better understand the disease and increase research outputs, improve disease surveillance, and inform the public health response. While there have been increasing calls, and sometimes requirements, to share data and samples collected, there has been little study of the ethics and governance of such sharing from the perspective of the professionals who may be involved.

In order to address this gap, this study aimed to examine perspectives of the professionals involved on the ethical issues, and the role of ethics governance, in relation to infectious disease data and sample sharing. This study has illustrated the ways in which data and sample sharing in the infectious disease context take place within a complex moral world. A number of practical ethical considerations emerged from the analysis.

In this chapter, following a brief summary of the main findings presented in Chapters 4 to 7, I reflect on and explore the implications of these findings in more detail and highlight the implications of the findings for policy and practice. The chapter ends by exploring some of the limitations of the study, and possible avenues for future research.

[17] Data and sample sharing are referred to as plural rather than singular, to reflect the finding that they are considered distinct by interviewees.

Brief summary of Chapters 4 to 7

In chapter 4, I focused on the ways in which views about different types of samples and data map on to views about what is acceptable and unacceptable to do with them. My analysis suggests that the categories of data, samples, and their sub-types have ethical implications. Some actions were considered permissible with some categories of data and samples but impermissible with others. This was often grounded in the degree to which the data or samples were seen to remain personal to individuals, thereby posing a risk of harm, and the potential utility of sharing for the infectious disease response. Similarly, 'sharing' was understood differently and was seen to incorporate a range of activities, including sharing between people who have a pre-existing relationship, and those who do not.

Chapter 5 took as its focus the roles of organisations and professions in outbreaks and infectious diseases. The analysis presented there showed that the plethora of organisations and professions involved in data and sample sharing in this context are perceived by stakeholders to have differing priorities, which are at times complementary, but also present ethical tensions. Sharing from low and middle-income countries (LMIC) was noted to bring particular considerations around justice – that the professionals and communities in LMICs receive a fair distribution of benefits.

Chapter 6 illustrated the ways in which the weight of ethical decision-making is felt at the individual level, with those involved in sharing discussing responsibilities to those from whom data and samples are collected, those who may benefit from outputs from data and sample sharing, and to other professionals. Claiming ownership and/or finding ways to control the data and sample sharing were put forward as ways to enable individuals to balance their perceived responsibilities.

Finally, in Chapter 7, the focus shifted to consent, community engagement, and governance as mechanisms with the potential to allow data and sample sharing to move forward in a way that was cognisant of the responsibilities identified in Chapters 5 and 6, but each was noted to also

have limitations. Developing trusted relationships was discussed by some as another way to enable data and sample sharing, but again, noted to be challenging to achieve and maintain, particularly when sharing with unknown others, or in rapidly evolving emergency situations.

In the following section (8.2) I explore the implications of these themes, and relate them to the wider literature, and policy and guidance documents.

8.2. Examination of thesis findings

What these findings show as a whole is that data and sample sharing in the infectious disease context are multi-faceted, and need to be understood as both constituting and taking place within the context of a complex moral world with different groups of people, with different norms, expectations, and ethics traditions, interacting around the purpose of data and sample sharing. There was broad agreement among interviewees in this study that data and sample sharing were important to respond to infectious diseases. Some interviewees considered the sharing of data, and the optimal use of samples, to be a moral obligation.

There were a number of differences between interviewee perspectives with respect to certain aspects of data and sample sharing. These differences are manifest both at the organisational and individual level, impacting upon judgements about when sharing is appropriate, how it ought to be done, with whom, and what ought to be in place in order to enable sharing. This is compounded by the potentially urgent nature of sharing in the context of infectious diseases emergencies that adds the additional dimension of speed into the decision-making.

I now move on to explore the ethical issues from the findings considering: the object – the data and samples (section 8.2.1); the differing priorities of organisations (section 8.2.2); the sense of personal responsibility of individuals (section 8.2.3); and the role and limits of some of the

proposed solutions (section 8.2.4). In addition to drawing upon the analysis in Chapters 4 to 7, cross-linkages between the chapters are also explored, as are connections with the literature and policy documents discussed in Chapter 2, along with literature, policies and guidance published subsequent to that. In regard to this latter point, one guidance document which is particularly relevant to this thesis is the Nuffield Council on Bioethics report on "Research in global health emergencies" in January 2020[18], which includes a chapter on data and sample sharing (143), and will be referenced at several points during the discussion as the "NCB report".

8.2.1. Delineating the objects, the way they are shared, and the disease

8.2.1.1. Ethical significance of the types of data and samples

The type of the data and samples being shared in outbreaks and in infectious disease research more broadly have ethical significance for people undertaking sharing, in particular around the potential benefits and risks related to sharing. As a result, interviewees tended to make distinctions between what is and is not permissible by reference to different types of data and sample types.

The distinction between data and samples

One key finding arising out of my analysis is that samples were generally seen as distinct from data in important respects by most interviewees in this study (Chapter 4 section 4.1.1). This contrasts with the position in certain policy and guidance documents. For example, the Médecins Sans Frontières (MSF) and International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) data and sample sharing policies discuss both data and sample sharing together (92, 115). The perspectives offered by interviewees in this study suggest that while there may be some similarities between data and sample sharing, ethics guidance relating to sharing them should not

[18] Some preliminary findings from this study were in fact presented as a part of an expert witness workshop during the development of the NCB report.

discuss samples as if they are a sub-type of data, to allow the distinct considerations to be more fully discussed.

Sensitivity of some types of data and samples

Some types of data and samples were considered to be particularly culturally sensitive. This echoes some of the literature, with some papers noting, for example, the particular sensitivity of certain types of samples. Blood samples are often discussed as being particularly sensitive within parts of Africa for various reasons, including the risk that withdrawing blood will be painful or result in ill-health, that the blood will be sold, or used for witchcraft (26, 144, 145). Sensitivities of different sample types have also been noted in a study in Scotland, where the study participants considered certain samples from inside the body such as blood, urine, faeces and semen to be more sensitive than samples such as saliva and nails (35). In keeping with the literature, some interviewees also discussed the particular sensitivity of genetic data, because it was seen to have the potential to reveal information about their own or others' predisposition to disease (146-148). Individual-level data, and data collected through research, were also considered to potentially provide more detailed information and therefore require greater protections (although it will depend on the specifics of the particular dataset). This reinforces the idea that 'data' or 'samples' are not one thing but are terms that cover different things which have different sensitivities, and these sensitivities can vary across different socio-cultural contexts.

One type of data that was not considered sensitive by interviewees in this study is aggregate data. The NCB report, however, discusses the risks of population-level stigma from data sharing (143 p.192). The report suggests that there is a potential risk even where data are aggregated or robustly anonymised, as information could be revealed about the population which is considered stigmatising such as a particular group having an increased risk of disease (48). The fact that this was not noted by many interviewees in this study may reflect limited awareness about this possibility.

The physical nature of data

One distinction commonly made between data and samples focuses on the physical nature of the samples, and the potential need to export them to share them. Some studies note concerns about where samples may be exported to, and a potential loss of control (26, 149). A surprising finding of this study is that some interviewees also discussed the physical storage of the data, in particular the location of servers, as ethically significant (Chapter 6 section 6.2.2.2). This related to a sense of retaining ownership and control of the data. It also seemed to have a more symbolic meaning, relating to concerns that the possibility of 'moving' data from the region was reminiscent of colonialism. This is an issue that was not raised in the literature reviewed in chapter 2. There is some literature on research ethics in relation to cloud computing, which discusses potential issues with data security, as data access may be granted to others by cloud computing providers (150). Data may also be spread over multiple locations, making them difficult to track (150). The implications of this for sharing data as a part of the infectious disease response need to be better understood.

Distinctions between data types

This study shows there are other potentially ethically relevant distinctions between data types that have received relatively little attention in the published literature. One such distinction, that would benefit from further unpacking in a future study, is between qualitative and quantitative data, with the few interviewees discussing this unsure, but debating whether there may be a risk for qualitative data of more personal information being shared (see Chapter 4 section 4.2.3.3). While the sharing of qualitative research data in general terms has been discussed in some of the literature (for example (151, 152)), it has not been discussed much in the context of infectious diseases responses, where data may be gathered and shared more rapidly.

New types of data also continue to be increasingly discussed in this context. The Covid-19 pandemic, ongoing at the time of writing, has seen significant discussion on the use of mobile phone data, and the use of apps in the response (153-155). These areas were not discussed much in the interviews I conducted, possibly because of the selection of people resulted in few of the interviewees being involved with the development of technology such as contract tracing apps, but also because this is a new and rapidly evolving area. As a result, a typology of data and samples is something that will continue to evolve and change over time.

8.2.1.2. The ethical significance of the difference between sharing data and samples with known and unknown people

Who the object (data or samples) was shared with was also seen as ethically relevant. Interviewees who were primarily data or sample contributors did, in general, seem more comfortable with sharing with those with whom they had an existing relationship.

This distinction between sharing with people with whom data or sample contributors have pre-existing relationships was often described in terms of trust, although the word trust was used in different ways by those discussing it in this study. What it seemed to imply, in general, was a way of referring to and understanding situations in which data or samples would be taken care of and used for good reasons (to advance science and benefit people, while being respectful to those from whom they had been collected). The expressed desire of those I spoke to, to share only with known people in established collaborations, is consistent with the findings of other studies on data and sample sharing, and global health collaborations more broadly (22, 117).

The difficulty this presents, particularly in unexpected situations such as epidemics, is that groups may have to work with collaborators who are not previously known to them. Restricting sharing to those with whom people have previous experience of collaboration would be in conflict with the practice and spirit of sharing promoted by WHO, funders, and journals. There is also

discussion in the global health bioethics literature about the difficulties using models of trust, in general, in research collaborations given unequal power relations between those in high-income countries (HIC) and LMICs (156). This provides a further challenge to relying on trusted relationships for data and sample sharing during international infectious diseases responses, and puts additional emphasis on governance arrangements which are discussed further below.

Where data sharing takes place to a repository, rather than with known collaborators, a small number of interviewees used the verb 'give' instead of 'share' to describe this. The term 'sharing' was contested by some of my interviewees due to a perceived lack of reciprocity (Chapter 4, section 4.2.2). There were some concerns around the moral baggage associated with the word sharing - giving the appearance that it is an unqualified good. Where sharing occurred through repositories, there was clear support to manage access to the data and samples.

8.2.1.3. The ethical significance of different disease contexts

In addition to the type of data or sample being shared, and who it is shared with, my research has shown the relevance of the disease to which the data or sample relates. The analysis found distinct ethical considerations are generated by different disease contexts in a way that impacts on data and sample sharing practices. Some of this was grounded in the fact that some infectious diseases such as Visceral Leishmaniasis disproportionately affect poorer people, who may be more vulnerable to exploitation, or conversely may benefit more from data and sample sharing as less relevant research tends to be undertaken on diseases that disproportionately affect them. These are issues that are not specific to infectious diseases. Issues raised that are more specific to infectious diseases were: firstly, and most relevant when contrasting outbreaks with non-outbreak situations, the potential urgency with which sharing may need to take place; and secondly, the stigma associated with the type of infection and the way it is transmitted.

The urgency of the disease response

The potential urgency with which data or samples may need to be shared was mainly discussed by interviewees in the context of outbreaks. However, some did feel that rapid sharing could have benefits for endemic diseases, as it could still prevent infection (Chapter 4 section 4.3.1). The urgency of data and sample sharing during outbreaks is discussed frequently in the literature. As noted in Chapter 2 the emergency context is often discussed as one in which there is a strong obligation to share (9, 37, 50, 109, 110). Some of those interviewed for this study did, conversely, consider that it may be more appropriate to restrict data sharing to more specific groups of people during an outbreak, due to the chaos and fear, and potential for misinterpretation of the data.

A further issue that was discussed in relation to emergencies was the potential impetus to act outside of usual norms.[19] This is a live issue – for example, during the Covid-19 pandemic there has been increased sharing of identifiable health data between organisations relevant to the Covid-19 response in England (157). Some interviewees felt that acting outside of usual norms was necessary to effectively respond and save lives. Others, however, were concerned that allowing people to act outside of usual norms could open the door to unethical practice. One example given of such unethical practice was the exportation of samples without the involvement of the originating government (Chapter 7 section 7.3.2.2) – an issue that has been reported more widely in relation to the export of Ebola samples from West Africa (158, 159). This risk is discussed in the NCB report – the issue that in emergencies there "can be increased pressure to cut corners" (143 p.82). It is clear from the interviews conducted for this study that there are divergent perspectives on the implications of emergencies for data and sample sharing practice, particularly in relation to how, if at all, they ought to change practice.

[19] Although not discussed by interviewees, the threshold at which something can be called an 'emergency' is of clear relevance here, and examined further in the NCB report.

The relevance of disease stigma

The stigma associated with certain diseases was discussed as relevant in situations in which individuals are either seen as sources of disease, or stigmatised for behaviours that are associated with susceptibility to the disease (Chapter 4 section 4.3.2). This was noted by some interviewees as a reason to protect certain data or samples more, particularly with respect to privacy, and maintaining confidentiality. This was particularly discussed in relation to sexually transmitted infections and blood borne viruses, including HIV. Where data and samples are being shared internationally, sensitivities will differ by context, and indeed the legal implications of being identified will also different across jurisdictions. With respect to HIV, for example, there are different laws regarding criminalisation of behaviours[20] (160), which would pose risks to individuals if there were any way of identifying them as a consequence of widespread data sharing. Stigmatising diseases that are part of an outbreak may cause particular tensions as there are competing considerations acting both for and against widespread sharing.

8.2.1.4. Thinking about the object, the way they are shared, and the disease

Bringing these points together – the object, the mode of sharing, and the disease context – it is clear that 'data and sample sharing in the infectious disease context' in fact relates to multiple different activities. The following list is an example of different activities that could be considered to be 'data and sample sharing in the infectious disease context' (not an exhaustive list):

- Sharing aggregate data openly on a widely accessible website.
- Sharing individual-level identifiable data without consent through a repository.
- Sharing routinely collected samples from an outbreak through a governed biobank.
- Sharing identifiable data related to an endemic disease within a close collaboration.
- Sharing qualitative data from a study about a stigmatised disease via a repository.

[20] These behaviours include same-sex sexual activity, sex work, and injecting drug use. As noted in Chapter 6, there are also some countries that criminalise the transmission of HIV.

The discussion so far has shown that when examining the ethics, the multi-faceted complexity of what is being undertaken needs to be considered, as each of the activities noted above has some distinct considerations. In what follows I present the findings on the differing organisational and professional priorities.

8.2.2 Organisations and professions, and their differing priorities

It is clear from the interviews conducted for this study that infectious disease research and response is increasingly an activity that calls for the establishment of very large-scale complex, multidisciplinary networks. The presence of a wide range of actors in data and sample sharing activities that emerges from my analysis echoes to some extent the literature on this topic (see Chapter 2 section 2.3.1.1). What this study adds is an empirical account of the varied practical and ethical implications of the involvement of this range of actors, and the ethical tensions that arise due to their differing priorities.

The infectious disease context adds a potential layer of uncertainty as organisations come together who may not have previous relationships (depending on the disease in question) and/or different governments are primarily involved from one situation to the next. This has been a driver for the development of preparedness networks such as the African coaLition for Epidemic Research, Response and Training (ALERRT), which describes itself as "a multi-disciplinary consortium building a patient-centered clinical research network to respond to epidemics across sub-Saharan Africa" (161). ALERRT includes institutions based in Africa and in Europe.

8.2.2.1. The range of organisations, and their roles

The WHO was seen as a leader for data and sample sharing for infectious diseases, and a locus of knowledge sharing, but also an organisation which faces difficulties fulfilling this role. A small number of interviewees felt that the WHO was slow to act, and insufficiently engaged with those convening platforms (Chapter 5 section 5.1.1) – a charge that has also been levelled more broadly

in the literature. In particular, during and after the Ebola epidemic in West Africa (2014-16) the WHO was criticised for its slow response by organisations such as MSF (162), and the Independent Panel on the Global Response to Ebola (163), with the latter noting that the initial phase of the outbreak "underscored inadequate arrangements between governments and WHO to share, validate, and respond robustly to information on outbreaks." Notwithstanding these concerns, interviewees in this study saw the WHO as a key organisation for data and sample sharing.

A number of other organisational types were discussed by the interviewees – government organisations which include public health functions, academic organisations, healthcare organisations, non-governmental organisations (NGO), commercial organisations, funders and journals (see Chapter 5 Figure 5.1). The role of public health, and the priorities that public health professionals were perceived to have – to share in order to better respond to infectious disease threats and, potentially, to save lives, was generally uncontroversial among those interviewed in this study. This was particularly the case for outbreak situations, where there was acute recognition of the interrelation between data and sample sharing, and the infectious disease response. This is reminiscent of the idea of solidarity discussed by Langat et al. in their paper on data sharing in public health emergencies (50). The paper discusses the important role data sharing has in preventing disease, and the need for a common response in the face of a common threat (50).

At times, however, the public health priorities appeared to be conflict with other priorities, which is a source of ethical tensions. This is discussed further below.

8.2.2.2. Differing priorities of different organisations and professions

While interviewees discussed a wide range of organisations as relevant to, and involved in, data and sample sharing, it was noteworthy that two types of organisation were particularly prominent in this study – public health and academia, possibly because many of those interviewed in this study were based in these types of organisations.

The partially overlapping priorities of academia and public health

The main complementarity of priorities was between the public health priority to share data and samples in order to better understand the disease and ultimately save lives and reduce morbidity among the population, on the one hand, and the academic priority of doing research, on the other. These are interlinked priorities, as better understanding of the disease will be related to research on the disease in question, and the development of knowledge (from a variety of disciplines) and innovations – such as diagnostics, treatment and vaccines – that better enable a public health response. Indeed, it became clear in the comments of my interviewees that there is often a potential blurred boundary between public health practice and research activities when responding to infectious diseases, as was also highlighted in the NCB report (143 p.21-22)..

Ethical tensions in the priorities of academia and public health

While there are clear complementarities between public health and academic priorities, there are also tensions. As in the literature (Chapter 2 section 2.3.3.2) some considered that the pursuit of academic recognition may hamper the potential of both the public health and academic (scientific) gains of data and sample sharing to be realised, as contributors may be reticent to share and lose recognition. The tension was seen as likely to be more acute where there are resource discrepancies, with the risk that those collecting the data or samples in LMICs do not get full recognition for their work, as those from institutions with more resources and who are able to analyse data or samples more quickly. Given these concerns, the NCB report discusses protections for local researchers, such as privileged access for a period of time to enable work on the data and/or samples (143 p.210-211). The report notes that involvement of local researchers can help prioritise research appropriate to the context, and help build capacity. Platforms such as the Infectious Diseases Data Observatory (IDDO) explicitly aim to include researchers from affected countries to analyse the collated data (164).

There is an apparent ethical tension in relation to emergencies, in particular where data and sample sharing may need to take place more rapidly and there may not be time to give such privileged access. On the one hand there is an issue of justice at stake, as those who put in the work collecting the data or samples, potentially from LMICs, may lose out on recognition. On the other hand, there may be an ethical (consequentialist) imperative for sharing to happen so that data and samples can be analysed as quickly as possible to aid the public health response. In the light of this, the NCB report notes that in emergencies protections for LMIC researchers may be temporarily unjustifiable in the face of public health needs (143 p.210-211).

However, the analysis presented in earlier chapters suggests that there are different perspectives on this apparent ethical tension. From the perspective of the data and sample contributors themselves there may not be a simple delineation of equity between professionals on the one hand, and meeting beneficiary interests on the other. The interests appear to be interwoven. Data and sample contributors were not necessarily focused on academic recognition. They were also concerned about how the data or samples would go on to be used, and whether they would be used for well-conducted studies. While there has, rightly, been focus on looking to address discrepancies in academic recognition, this study shows that data and sample contributors also perceive their involvement to be in the interests of preventing misuse and enabling good science.

Ethical concerns about commercial priorities

Another notable finding in this study was the mixed feelings felt by some of the interviewees around the potential priorities of commercial organisations involved in data and sample sharing (Chapter 5 sections 5.1.2.4 and 5.2.5).[21] Pharmaceutical companies were exclusively discussed by those interviewed in those study, but there are other relevant commercial organisations such as technology companies who may be involved in large-scale data analyses, and the development of

[21] Of note, no one from a commercial organisation was interviewed in this study.

contact tracing apps (154). Some who had more experience involving pharmaceutical companies reported that they could be valuable partners in the sharing process, and they recognised their potentially important role in developing innovations. The main concern related to the organisations profiting from the development of innovations, and that the countries in which the data and samples had been collected would not be able to afford those innovations. The analysis shows that it was not straightforwardly the case that commercial organisations were perceived to be a problem. Interviewees discussed a mix of what were perceived to be 'good' and 'bad' reasons for them to be involved in sharing. While the priority of commercial organisations to contribute to research and develop innovations was seen as 'good', the motivation to derive profit was seen as 'bad'.

This resonates with the findings of a survey developed by IPSOS-MORI for the Wellcome Trust, which studied the views of a very different group of people. The survey was completed by members of the general public on commercial access to health data in the UK. It showed that people considered the balance between public benefit and private profit to be important, with concerns in particular where there was a perception the data would primarily be used for private profit (165). It is interesting to note a similarity with perspectives offered by interviewees in this study.

8.2.2.3. The power of funders and journals

A further finding emerging from the analysis was the perceived dominance of certain types of institutions – particularly funders and journals – in mandating data sharing. My research showed a mixed reception to these requirements, with some interviewees welcoming this move, but others highlighting a number of concerns – such as the lack of attentiveness given to potential concerns of data contributors, the lack of funding given to prepare data for sharing, and the power of funders augmenting the power of those from HICs, as funders are predominantly from HICs (Chapter 5 section 5.1.3.1). These are important findings, as the perspectives of those engaged in

sharing with respect to these issues have not previously been studied –it shows that these requirements are not straightforward enablers for sharing, but may in some circumstances be a seen as an imposition.

This could lead to different issues. One echoes a concern identified in a paper by Merson et al. on data dumpsters, which highlights the risk that sharing takes place to tick a box, in a way that does not actually facilitate the use of data sharing (for example lacking metadata or data dictionaries) (80). I believe this is ethically relevant – if data and sample sharing have the potential to improve health and save lives, it needs to be done in a way that is conducive to that outcome. This suggests a need for funders to be more specific about what is being asked of researchers as a part of 'data sharing' and to ensure that repositories are appropriately funded to make them useful.

An additional issue raised by the requirement to share data is that of power of those based in HICs to set the norms. This looks to have some grounding. The Global Research Collaboration for Infectious Disease Preparedness (GLOPID-R), a collaboration of international funders for infectious diseases, has produced a set of principles, and a roadmap, for data sharing in public health emergencies. These were developed during the time this study was taking place. Examining the GLOPID-R working group for data sharing at the time of writing (August 2020), of the 25 participants, all but one are either from organisations in HICs, or from supra-national organisations such as the WHO (166). There is one person based in an LMIC organisation (in South Africa). The GLOPD-R data sharing working group website does, however, also note it is seeking more diversity including wider geographical representation (167).

The importance of funders and journals meaningfully engaging with data contributors across different locations is underscored by the arguments above, to enable requirements to be sensitive to the varied considerations relevant to data sharing.

8.2.3. Ethics at the individual level: A sense of personal responsibility

Another facet of the findings of this study is that ethical considerations and moral dilemmas were felt very personally by those engaged in sharing. This often took the form of a strong sense of responsibility. The analysis presented in Chapter 6 showed there to be three broad groups of people that interviewees considered themselves responsible towards:

- people from whom data and samples were collected;
- people whose health may benefit from work on shared data and samples;
- and other professionals who were also engaged in sharing (generally discussed as ensuring data or sample contributors are recognised).

Reflecting on the analysis presented in Chapters 4 to 7 as a whole it is apparent that there is a potential tension between meeting these responsibilities, and that different interviewees drew the boundaries differently with respect to which group's interests they would prioritise. This appeared to be related to their 'proximity' to the people that data and samples had been collected from. Researchers and healthcare workers were therefore more inclined towards greater protections for the people from that group. Meeting different responsibilities entails different levels of constraints on sharing. Ownership and/or control were deployed as ways of meeting perceived responsibilities to others by enabling boundaries to be put in place for how sharing takes place.

8.2.3.1. Responsibilities to the wider population

As noted above, it was largely uncontroversial among this group of interviewees that data and sample sharing were of importance for public health, when sharing with people who had the knowledge and skills to use the data in ways that produced good analyses. Some felt that there was a moral obligation on data contributors to share data, and sample contributors to use samples optimally (although as noted previously the sharing of samples with others was not considered to be such a straightforward obligation). Given the potential speed at which sharing may need to take

place in a rapidly unfolding emergency, timeliness of data and sample sharing was considered ethically important by some interviewees.

While the public health benefits of sharing were uncontroversial, there were concerns about which 'public' stood to benefit from the data and sample sharing. The distribution of benefits from data and sample sharing is a well-recognised concern. As discussed in Chapter 2 section 2.3.5.4 in relation to influenza vaccines (84), there was concern that research conducted using shared data and samples will primarily benefit HICs who are able to purchase any innovations developed from shared samples, or drugs and diagnostics shown to be effective through work on shared data. At the individual level, this translated to a concern about being part of an unfair system.

8.2.3.2. Ethical tensions between meeting responsibilities to the people data and samples have been collected from, and those who might benefit from their use

There was a clear ethical tension notable in this study, between meeting responsibilities to those from whom data and samples were collected, and to the people whose health may benefit from data and sample sharing. Meeting responsibilities to the wider population was generally seen as requiring people to share rapidly (albeit with consideration of who then benefits from this rapid sharing), and protecting the people that data and samples were collected from was perceived to act as a constraint. While the difficulty of balancing individual and population interests is discussed in the literature (48, 168), it is often in an abstract sense. This study shows the degree to which some of the individuals who are involved in data or sample sharing feel this to be a personal moral dilemma, and find difficulty balancing these interests. Some interviewees felt that the data or samples were in fact a part of the patient. They were acutely aware of their link to patients and research participants, and as a result they felt personally responsible for taking care of them. Some of this relates back to the areas discussed under 'object' – what type of data or samples are being shared, and how sharing is being done – and whether data being shared are at the individual level, for example.

Consistent with the literature in Chapter 2, interviewees raised concerns about privacy, and whether benefits would come back to communities that the data and samples were collected from. One particular concern discussed in this study is that for some infectious diseases with a short clinical illness, and subsequent immunity (or death) the people from whom the data or samples were collected would not benefit as a result of them being shared. This raises a question of whether some form of benefit sharing is due, where benefits are not expected to those individuals from the knowledge or innovations derived from sharing, or whether individuals should expect that such sharing will take place without any returning benefits to them personally, as a part of our global collective responsibility to respond to infectious diseases. This is a question more relevant to sharing routinely collected data or samples, which would not fall under discussions of post-study obligations.

8.2.3.3. Ownership and control, and meeting responsibilities

The analysis presented in Chapter 6 highlighted the use of the concept of 'control' as a tool for people, particularly data and sample contributors, to talk about ways in which to meet their responsibilities, while (to some extent) engage in/be seen to engage in sharing. Control mechanisms discussed included claiming ownership over the data and samples, or acting as gatekeepers for the sharing of data and samples. Some policy documents, such as the ISARIC data and sample sharing policy, explicitly discuss data and sample contributors as owners (115)

The findings show that interviewees had other motivations for constraining or controlling sharing, beyond the authorship concerns which are often cited in the literature. These ownership and control claims look to be methods of aligning practice with perceived responsibilities to the groups outlined above. Although control mechanisms were also perceived to be used for other reasons, beyond enabling people to meet their responsibilities, such as concern about data being scrutinised and remaining connected to the data or samples.

Some authors argue that publicly funded data are a common good and should be used to improve the health of the public (50). Others note that discussion of ownership is unhelpful, and that stewardship or custodianship models better reflect the need for ongoing care and management of data and sample repositories, to ensure they are used in a way that is scientifically ethically appropriate (20, 169, 170). While both of these points give reasons to move away from discussion of ownership and control, it appears from the stakeholders interviewed for this study that it is important to continue to explore the motivations behind data and sample contributors seeking ongoing ownership and control. This could deepen understanding of the issues, including ethical issues, they may be concerned about in that specific instance of data and sample sharing.

8.2.4. Enabling ethical data and sample sharing

The previous sections, focused on the analysis in Chapters 4, 5, and 6, have shown that there are a number of tensions with respect to ethical practice. This section, which focuses mostly on the implications of the analysis presented in Chapter 7, discusses the different ways put forward to enable data and sample sharing to take place. I have collectively termed these 'solutions'. It is clear from the findings in Chapter 7 that all of these 'solutions' were seen to be a potentially useful to enable data and sample sharing that addressed some ethical considerations, but also had limitations.

8.2.4.1. Consent and community engagement

Consent, and community engagement were perceived to play a role in ensuring sharing is respectful of the wishes of the people from whom data and samples were collected, either individually or collectively. If consent was there, or the community has been engaged, this was seen as giving permission, to some extent, to share, and alleviated some other concerns that those contributing data or samples may have with respect to protecting those from whom data or samples are collected.

It is notable that while an initial almost textbook discussion of the importance of consent, and the potential role of broad consent, took place, interviewees also noted a number of potential issues with obtaining consent in practice. Firstly, some doubted how much would be understood about data and sample sharing by those from whom consent was being requested, for example because people might be very unwell at the time. This raises questions of capacity to consent. Secondly, in keeping with the wider global health bioethics literature (171-173), some discussed the need for additional authorisation from community leaders, in addition to individual level consent. Thirdly, a specific issue with certain types of infectious diseases was noted. Where personal protective equipment was required to be worn, and this could only be done for time limited period, there was a question about whether obtaining consent for data and sample sharing was a good use of this time, given it would by definition displace some clinical care (Chapter 7 section 7.1.1.3). This brings to the fore a larger question – what priority should be given to discussions about data and sample sharing with individuals and communities, when there are a number of pressing issues that need to be discussed during an epidemic?

There seemed to be much less of a consensus about the role of community engagement, and how it related to making sharing ethical. Some considered it to be a way of informing the community that the data and samples originated from that sharing was taking place, while others discussed it as a way of gaining collective permission. Some formal roles of the community are discussed in the research ethics literature, such as the use of community advisory boards to provide representation (36, 174, 175). Within the context of a rapid infectious disease response, however, there were noted to be competing priorities beyond data and sample sharing when engaging the community. Furthermore, as noted by Wright and Sheather, in a global public health emergency there are not likely to be community advisory boards in place (176). Echoing the literature, some interviewees also voiced concerns about who could effectively represent the diversity of a collective group of people, if community representatives are sought (177).

8.2.4.2. Committees and agreements

Governance mechanisms, such as research ethics committees (REC), data sharing agreements (DSA), material transfer agreements (MTA) and access committees appear to operate very much at the interface of different professional groups, with roles at different stages of the sharing process (Chapter 7 section 7.3). MTAs are also legal documents. However, in this study they, alongside the other governance mechanisms, were discussed in relation to how they are used in relation to ethical practice, rather than their role in relation to the law.

RECs were discussed as a permission-giving authority whose involvement was considered most relevant prior to sharing taking place – and were discussed in relation to data or samples that were collected through research and/or being used for research purposes. They may not, therefore, be involved in relation to sharing of routinely collected data or samples for surveillance or other public health purposes. RECs were also seen as a way of incorporating local perspectives into the appropriateness of sharing – that is, perspectives from the place from which data or samples were collected. It was notable that there were differing levels of confidence in RECs – and some, particularly those based in HICs, considered that they were not necessarily aware of all of the issues, or did not have sufficient time for a robust ethics review in an emergency (Chapter 7 section 7.1.3.2). The need for ethical preparedness has been increasingly recognised following infectious diseases outbreaks such as the Ebola outbreak in West Africa and Zika outbreak in Latin America. The WHO and ALERRT convened a workshop and have developed recommendations for RECs including in relation to data, sample and benefit sharing (178). The report also highlighted the need for more capacity strengthening in relation to some of the arising issues. The perspectives offered in this study support the importance of such ethical preparedness.

The role of agreements, such as MTAs and DSAs, was considered to be at the point of sharing the data and samples between groups or to a repository, laying out the expectations between the

different parties involved in data or sample sharing. The access committees were discussed as arbiters of whether the person or group requesting use of the data or samples from a repository was making an acceptable request – that they were qualified, and that the use was a good use of the data or samples – with being qualified and a good use both being areas to define. These processes are seen as potential other routes to help develop ethical practice. Agreements, for example, can delineate certain terms of use and discussion of issues such as the distribution of benefits, and claiming intellectual property. Access committees, if effective, could stop sample data and samples being shared where the request is considered inappropriate. An inappropriate request may be one where the planned study had a poor methodology that could then lead to low quality information being disseminated, or where the use would be derogatory to particular groups (although this would be a matter of judgement). They are also a way of prioritising the use of finite samples.

However, it is notable that a small number of interviewees in this study seemed unclear about whether a process such as a committee or agreement was adding value, or what their purpose was, among the other governance tools that are also relevant when sharing data or samples (Chapter 7 section 7.3). There was also a concern that processes, if cumbersome, in general could delay data and sample sharing thereby impacting on the achieving rapid sharing for public health purposes.

8.3 Perspectives on what constitutes 'ethical practice' during infectious disease responses

Following from the examination of the findings above, this section focuses on taking the findings as a whole to examine what is considered to constitute ethical practice of data and sample sharing for infectious disease responses by those interviewed in this study. The findings above show that interviewees raised a number of well-recognised ethical considerations, but also others

considerations that are often not discussed as 'ethical' in nature. Some of these considerations were impacted by whether or not the infectious disease in question was a part of an outbreak.

8.3.1. Well-recognised ethical considerations in data and sample sharing

There are a number of ethical considerations discussed by interviewees that are well-recognised in the literature and guidance, as discussed in Chapter 2 of the thesis. Examples include: protecting the interests of the people data and samples were collected from; considering the role of privacy and confidentiality; the importance of benefit sharing, and the wider distribution of benefits; the distribution of academic recognition, and capacity building; and the role of consent and community engagement, and RECs for ethical practice.

It is notable that interviewees had different perspectives in relation to these considerations, and one factor in these differences appeared to be their proximity to the people that data and samples had been collected from – with those who were involved in research or healthcare being more inclined to advocate greater protections for that group. A further notable feature of the findings is that for some interviewees things were more straightforward and there was a dominant value, for example, an overriding obligation to share data (Chapter 6 section 6.1.2), or to only share if there was consent (Chapter 7 section 7.1.1.1). Other interviewees seemed to be weighing up factors, and to feel a greater tension about the best course of action.

8.3.2. Factors influencing ethically permissibility of sharing

With regard to what constitutes ethical practice, findings from this study show that a number of factors seem to promote or constrain sharing, and interviewees made different judgements about the extent to which the factors do so. Those emerging from the analysis are summarised here:

- the type of infectious disease, and in particular whether it is part of an outbreak or not;
- the object – whether it is 'data' or 'samples' and the different data and sample sub-types;

- whether sharing is taking place with people who are known or unknown to the contributors, and resultant trust and confidence in the sharing system;
- whether the sharing is a part of research or the response, or both;
- the organisations and professional groups involved, and their suspected priorities, which included a notion of 'good' and 'bad' priorities when engaged in sharing, such as those identified for commercial organisations above;
- whether data or samples are collected and shared in HIC or LMIC settings;
- the extent to which individual and population interests are being met;
- and whether certain 'solutions' were in place such as consent, community engagement, and ethics approval.

These factors are themselves not straightforward to delineate – it is notable from the discussion in previous chapters that it is difficult to differentiate 'research' from 'response', or fully unpick varied organisational priorities. The factors also interact with each other. For example, when thinking about the type of data or samples, interviewees had some concerns about whether something was anonymised enough for sharing – the sense here being that the data or samples needed to be more robustly anonymised (Chapter 4 section 4.2.3.5). However, the context of an outbreak changed this for some, who thought the emergency response justified the sharing of more identifiable information with those engaged in the public health response.

8.3.3. Considerations that may not be framed as 'ethical'

This study shows that, for those involved in infectious diseases research and the public health response, what is considered ethically relevant goes beyond much of the guidance, where the 'ethics' section is often concerned with consent, confidentiality and benefit sharing as noted in Chapter 2. The point of this section is not to refute the ethical importance of these areas, and those engaged in sharing in this study certainly echoed the literature that these are of clear relevance

to their work. Rather, the point is to highlight other areas that may not explicitly be considered to be 'ethical' in guidance, but clearly were seen as such by those engaged with sharing. Whether or not something is framed as ethical can then affect the extent to which it is actively weighed up when making judgements about ethical practice when sharing, either at an individual level, or by groups explicitly charged with providing a judgement such as RECs.

Other issues that were raised frequently in this study, which are noted in section 8.2 include: timeliness; quality of data; purpose of use; and the location of sample repositories and data servers. With the exception of the location of data servers, these issues are frequently discussed in guidance, particularly in the emergency context, but may be separated from the section on ethics.[22] For example, the GLOPID-R data sharing roadmap for emergencies – a publication from a group of international funders of high profile - identified has the following principles for sharing: Timely; Ethical; Accessible; Transparent; Equitable; Fairness; and Quality (179). Under the 'Ethical' principles the roadmap states:

> *Sharing of data must be done in accordance with applicable ethical and legal standards, ensuring respect for the privacy of individuals and the dignity of communities. This is essential for building the trust of the public. Additional attention should be given to respect for, and alignment with, cultural norms. This may include consultations with knowledge keepers, community members, local leaders/elders, and following appropriate protocols. Recognising the ethical importance of consent, informed consent models that allow for secondary use of data with conditions, known as broad consent, should be used to the maximum extent possible* (179).

The Chatham House ethical principles for data sharing states the following principles: Social Value; Respect; Justice; and Transparency – but do not specifically note issues such as timeliness (180). On the other hand, the MSF data sharing policy states that "MSF recognizes the ethical imperative it has to share its data openly, transparently and in a timely manner for the greater public health

[22] There are a large number of relevant guidance and policy documents, as outlined in Chapter 2, and also some that have been published subsequent to the review. It is difficult to generalise across all of these – the point here is that certain well-recognised sources of guidance aimed at the data and sample sharing community during infectious diseases responses have a narrower inclusion in relation to the 'ethical'.

good" (92 p.4). Timeliness is here noted to be a part of the ethical imperative, which mirrors the perception of several interviewees in this study.

A further area noted above that was discussed as ethical in character by some interviewees in this study (from sub-Saharan Africa), but is not much discussed in the literature or guidance, is the physical location of the data – that is the location of server(s) that the data are hosted on. This was related to a feeling that the data are being taken and used with a risk of loss of involvement or control. An area that requires further exploration is how this relates to ethical practice. Does the location of servers have ethical implications? If so, how does this fit within judgements of ethical practice?

A further question raised in this section is the motivation on the part of the interviewees to discuss these issues in a study about ethical practice. This is not something that can be fully answered by the analysis in this study. It may be the case that those interviewees raising these considerations genuinely believed them to be ethical in nature and want it to be included in judgements on ethical practice. It is also important to acknowledge that it is also possible that some interviewees believe that badging something as ethical is instrumentally valuable to reach other goals – that they want sharing to happen in a particular way (for examples, servers to be used from a particular location), and calling it ethical gives more weight to enable sharing to happen in that way. While this study does not answer this question, it was apparent that interviewees for the most part took the 'ethical' seriously, so I would not want to assume there is an ulterior motive.

8.3.4. Multiple perspectives on what constitutes ethical practice

As noted at the very beginning of this chapter, this study shows that those involved in data and sample sharing as a part of infectious disease responses are navigating a number of ethical considerations, and factors that are perceived to have ethical implications. When responding to an infectious disease, the sharing often takes place between a diverse set of individuals and

organisations who make different judgements about how sharing ought to take place for it to be considered ethical. This is complicated by a context that may be an emergency context, where there is increased pressure for rapid sharing for public health gains, while at the same time noting ongoing concerns about the protection of those that data and samples were collected from, and the distribution of benefits related to knowledge and innovations derived from data and sample sharing.

A key finding of this study is that what constitutes ethical practice is contested: firstly, in relation to what is ethically relevant, with some interviewees discussing factors that are not necessarily badged as 'ethical' as noted above; and secondly, in relation to how the factors identified in sections 8.3.2 ought to impact on practice.

The practical implication of this is that, even if one could claim that there is an objective take on what ethical practice is within a given instance of data and sample sharing, this study shows that it would be difficult to arrive at the answer as there are differing perspectives on this. At the organisational level, the findings show that some institutions have more power to influence, or even mandate, whether and how sharing is done – with funders and to a lesser extent journals being seen as particularly influential (Chapter 5 section 5.1.3). At the individual level, the findings show that the background of the individual, in particular their professional group, but also where they are from and/or based and their proximity to research participants or patients, looks to influence their judgements.

In the NCB report chapter on data and sample sharing, there was discussion of an 'ethical compass' as a set of three values that are widely accepted that can guide deliberations and action (143 p.76-101). These values are equal respect, helping reduce suffering, and fairness.[23] The report discussed

[23] These areas are defined as the following: "Equal respect: treating others as moral equals, including respecting their dignity, humanity and human rights; Helping reduce suffering: acting in accordance with fundamental duties, founded on solidarity, and humanity, to help those in need or suffering from

the wide range of frameworks and guidance related to research in emergencies, the plurality of individuals and organisations involved, and the challenging circumstances of an emergency. Given these circumstances, the compass was put forward as a way of guiding deliberations between those involved in the research, using those values. The approach of using the ethical compass also very clearly acknowledges the importance of being sensitive and attentive to a plurality of perspectives. It is notable that the importance of these values is not contested, and is very much reinforced by those interviewed for this study. However, if we take seriously the idea that perceptions of ethical practice relate to firstly, people's professional ethical norms and secondly, the proximity or distance from the people from whom data and samples had been collected, this may make the use of an ethical compass more challenging. Different professional groups may follow the norms of their profession when making judgements between these potentially competing values, rather than being able to make a balanced assessment.

8.4. Recognising divergent perspectives: the role of 'boundary objects'

Given that that ethical practice is contested, and groups looked to be following their differing professional norms, how can data or sample sharing take place in a way that is satisfactory to different stakeholders? This section takes a tentative step to offer progress answering this question, using the concept of 'boundary objects'.

The concept of 'boundary objects' was initially developed by Star and Griesemer, through research on the development of the Museum of Vertebrate Zoology at the University of California, Berkeley (181). They studied how different groups (called social worlds), such as professional scientists, amateur collectors, administrators, and trappers, were able to cooperate in the development of the

disease; and Fairness: including both duties of non-discrimination in the treatment of others, and of the equitable distribution of benefits and burdens."

museum as a site of scientific research, despite their divergent priorities and perspectives in relation to their involvement with the museum. The paper discussed how museum specimens, for example, were seen differently by different groups. Some saw them as objects of scientific enquiry, while others saw them as a way of preserving local flora and fauna, or as a source of income. Alongside standardisation of methods[24] Star and Griesemer discussed the role of these boundary objects in enabling cooperation despite these divergent priorities and perspectives. They define boundary objects as:

> … *objects which are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use* (181).

Star refrains from giving a strict response as to what could or could not be a boundary object, stating two features instead – it's scale, stating the concept "is most useful at the organizational level" and its scope, using the concepts in relation to groups of people, rather than in a generalised way (182).

There are a wide variety of potential boundary objects discussed in the literature, including repositories, objects such as maps, notes and documents, processes, and even more abstract ideas such as theories, or even a word or phrase (182). With respect to data and sample sharing, studies have conceptualised the biobank itself (18), the informed consent process for future uses of tissues (183) and the data management plan (DMP) (184) as a boundary object. What the use of the concept brings to those studies is an analysis of how these objects are flexible, and changing, in their use. It foregrounds how these objects enable communication and cooperation between different groups of people despite the fact that these groups come with differing conceptualisations about what the biobank, consent process, or DMP respectively are looking to

[24] Standardisation of data collection methods is relevant to data sharing. This was not, however, something that was discussed much by interviewees – perhaps as the focus of the study was explicitly on 'ethics'.

achieve. This looks to be of clear relevance to explore the findings of this study, as those involved in data and sample sharing need to agree on a way to share data or samples, despite divergent views on how to do so ethically. Furthermore, different group of people who may be unknown to each other may be coming together in a particular instance of data or sample sharing, leading to a changeable dynamic between sharing parties.

It is the conceptualisation of agreements between sharing parties as boundary objects that is most relevant to the findings of this study. Looking at the proposed 'solutions' discussed in Chapter 7, agreements are those that most explicitly act at the interface to the professionals directly involved in sharing. To some extent they seem to provide an object around which organisations, and the associated individuals, with different priorities, values and conceptions of what is ethically relevant can deliberate, adapt the agreement, and forge a way forward. While access committees also act at the point of sharing, their role is to deliberate on sharing, once data or samples are already in a repository, arbitrating on whether those applying to use the data or samples are making an appropriate (scientifically and ethically) application.

The findings presented in Chapter 7 show that for some, agreements provide a way of remaining connected with the data or samples, and retaining a sense of control over what happens to them. They also provide a way to discuss and agree upon specific issues such as the distribution on benefits and intellectual property claims (Chapter 7 section 7.3.2.1). This could help those involved in sharing feel, at least to some extent, that they are meeting their responsibilities to others, as well as protecting their own interests. It is possible that different groups would look at the same agreement, with one group believing that it has achieved the aim of enabling rapid sharing, whilst another believing that it has achieved the aim of ensuring those individuals that data and samples were collected from have been appropriately protected, for example. The agreement could thereby act at the interface of different individuals and organisations who weigh responsibilities to the

individuals that data and samples had been collected from, and the wider population who may benefit from the development of knowledge and innovations differently.

As noted in Chapter 7 some felt agreements did not add anything, beyond the requirements that are already in place (Chapter 7 section 7.3.2.2). Conceptualising agreements as boundary objects could be a way of clarifying their purpose and how they can add value in relation to ethical practice. It brings to the fore a recognition of the diversity of perspectives, and lack of consensus on what ethical practice is, alongside the fact that data and sample sharing happens in different contexts, with different factors at play that will impact upon the ethical considerations. This gives a reason to forge the agreement in an inclusive manner.

One specific example relevant to the infectious disease context is the WHO Research and Development (R&D) Blueprint sample Material Transfer Agreement. The WHO R&D Blueprint is described as "a global strategy and preparedness plan that allows the rapid activation of R&D activities during epidemics. Its aim is to fast-track the availability of effective tests, vaccines and medicines that can be used to save lives and avert large scale crisis" (139). It is focused on a list of diseases that are considered to be of epidemic potential, for which there is limited knowledge. As a part of the blueprint, a sample MTA is being developed through consultation, and is specifically intended as a flexible tool to be adapted for use by those involved in sample sharing for public health emergencies, particularly when sharing involves low-income countries. They state that the sample MTA aims to increase capacity to develop MTAs, and enable sample (and associated data) sharing to happen for response to public health emergencies, while protecting the interests of the sample contributors (185). There is an emphasis on benefit sharing. Such a document is intended to be general, and non-specific, and can then be adapted to the context, and so could serve as a boundary object between the organisations and/or individuals directly involved in sample sharing. It can provide a "shared… language for individuals to represent their knowledge" and potentially enable those involved in sharing to "jointly transform their knowledge" – two of the characteristics

of effective boundary objects discussed by Carlile following his ethnographic study of product development (186). It is possible to imagine a similar sample DSA, with considerations given for different types of data. Sources of guidance, such as the NCB report, alongside others discussed in Chapter 2 and in this chapter, that can provide additional guidance on the substantive ethical considerations when engaged in data and sample sharing as part of the infectious disease response.

Of course, there are practical realities that make the realisation of agreements as boundary objects difficult to achieve. The power of some organisations to have a greater influence on the content of an agreement, for example because they have more resources to negotiate, may mean that DSAs and MTAs do not satisfactorily achieve the aim of allowing cooperation without consensus, but instead predominantly impose one perspective on how the sharing ought to take place. The DSA and MTA would not then be acting as boundary objects. As stated in Star and Griesemer's paper:

> *The production of boundary objects is one means of satisfying…potentially conflicting sets of concerns. Other means include imperialist imposition of representations, coercion, silencing and fragmentation* (181).

The risk is where there are power imbalances, the agreements could serve as an imperialist imposition. This is a risk relevant to infectious disease data and sample sharing, where some groups are seen as being more powerful and influential than others.

Another factor influencing the realisation of agreements as boundary objects is the law. The content of the agreements will be constrained by the law within the jurisdictions across which sharing is taking place. This is an issue that is key to how data and samples can actually be shared, although closer examination of the law in relation to this topic is beyond the scope of the research aims for this thesis.

Finally, I am not making a claim that conceptualising agreements as boundary objects makes data or sample sharing ethical, nor that such a conceptualisation relates to a normative account of what practice ought to look like for data and sample sharing. The point is to recognise the diversity of

perspectives on the topic of ethical data and sample sharing in the infectious disease context, and using the idea of boundary objects looks to provide one way of thinking about how data and samples are, or could be, shared in practice, whilst recognising (and not stifling) this diversity.

8.5. Implications for policy and practice

As noted in section 8.3.1 above, the findings from this study lend additional support to the well-recognised importance of a number of ethical considerations. There are some additional considerations which have been less well examined in the literature, which I have summarised here:

1. The analysis shows that data sharing and sample sharing have some distinct considerations, which suggests that policies and guidance should not consider samples to be a sub-type of data. This would allow guidance to more fully outline the particular implications of sharing data, or sharing samples. There are specific considerations outlined for samples such as their finite nature leading to a need to prioritise their use, the connectedness of samples to the human body and the cultural significance of this, the need to tangibly move them to share them, and the potential to derive innovations that lead to IP claims. Similarly, for data, there were particular considerations around the ease with which they may be replicated and shared, leading to concerns about how to control them. The grouping of samples as a sub-type of data can mute the recognition and acknowledgement of these distinctions.
2. There is an argument to suggest the term 'sharing' should be avoided for some activities, as the term is overly broad. The idea that sharing is uploading data onto the internet in a way that is freely accessible was not supported by interviewees in this study, but it is one interpretation of the term. Without clarity on what sharing actually entails there is also a risk that people think they have done sharing, in keeping with the calls for sharing by funders and journals, without it being done in a way that realises the benefits, as noted by Merson et al. The use of specific language can guide those who may contribute data (this seemed to be less of an issue

for samples where, outside of collaborations, managed repositories seemed to be an assumed norm). For example, calls by funders for "rapid sharing" could be replaced with calls for "rapid transfer to a managed access repository" so that it is clear what is being requested. This clarity in language can also aid a clearer assessment of the ethical considerations associated with that particular activity by those involved in sharing, and those assessing the ethical implications such as RECs. (I have continued to use the term sharing for continuity over the thesis).

3. At the policy level, the findings show the importance of recognising that there are a range of organisations and professions involved in the infectious disease response, and a plurality of perspectives on what constitutes ethical practice. The interviews show that the nature of ethical practice is not settled, and some interviewees are concerned that certain organisations, such as funders, are in a more powerful position to set the norms. The importance of inclusive methods that enable diverse groups to participate in the development of policy and guidance is reinforced by this study. There is otherwise a risk that certain considerations are not recognised or addressed. This could help increase confidence in the system, if a wider range of issues are explicitly acknowledged.
4. Relatedly, at the practice level, as has been discussed above, the development of agreements can be an inclusive process by which sharing parties openly deliberate on their perspectives as to how sharing ought to take place.
5. Finally, the findings show that time-pressures during emergencies both create an impetus to share rapidly – as is often discussed – and a difficulty fulfilling the demands of usual good practice. The time pressures mean that there are competing priorities when managing patients or engaging the community that may limit the time and opportunity to obtain consent, or promote wider discussions about data and sample sharing. They may also mean that RECs are under pressure to rapidly review and provide a verdict on whether data or sample sharing ought to go ahead. This suggests guidance developed for these contexts needs to be explicitly cognisant of these challenges. For example, guidance documents will often assert the

importance of informed consent, and the benefits of broad consent, but there is more limited recognition of the practical constraints of obtaining any sort of consent in certain situations. Making these challenges explicit can provide an opportunity for those involved in sharing to proactively discuss the challenges ahead of time and consider whether it is possible for sharing to still take place in a way that is ethical (in that example, still respectful of the wishes of people that data or samples are collected from). This is also an area for further research, as will be discussed more in section 8.7.

8.6. Strengths and limitations of this study

This study has addressed an important research gap, providing perspectives on ethics and governance that are grounded in the practices and experiences of those involved in infectious diseases data and sample sharing. The analysis draws on the experiences of different types of professionals based in both HIC and LMIC settings, with experience of sharing different types of data and samples. This has allowed a wide-ranging examination of the topic, and unmasked important distinctions, competing priorities and ethical tensions when sharing data and samples. The inclusion of people with experience in relation to a range of infectious diseases, rather than focusing on one or two specific infectious diseases, has enabled exploration of the contrast between emergency and non-emergency settings, and different disease types. In addition, the use of interviews as the primary mode of data collection enabled a rich discussion and exploration of individual perspectives, and revealed the very personal values that weigh on people's minds in relation to data and sample sharing. Finally, the use of an initially broad set of questions during the interview enabled me to develop an overview of the experiences of the interviewee, in which the discussion of ethical issues was then embedded. This, I believe, has enabled an examination of the topic that is clearly grounded in practice.

There are also limitations to this study, which are important to recognise when considering the findings. Firstly, this study was explicitly focused on ethics and governance, and interviewees were aware that I am a student in a bioethics research unit. This may have led some interviewees to focus on discussing what they considered to be ethically relevant issues. As a result there is the risks that some issues I would have considered ethically relevant may not have been raised as the interviewees did not consider them to be ethical in nature.

Secondly, purposively selecting interviewees through the IDDO network enabled me to recruit people who were clearly relevant to the study. A limitation is that this may have led interviewees to present their views to me differently than they would have done otherwise – for example because they considered me associated with IDDO as I am also from the University of Oxford. Although there is no way to fully overcome this issue, I did attend meetings where the platforms were discussed and some of the interviewees were in attendance, which at least gave the opportunity to note discrepancies.

Thirdly, I made a choice to move away from a focus on examining data and sample sharing in relation to specific diseases to a wider inclusion of people with experience with a range of infectious diseases and different modes of sharing. This change was made to recognise the reality that interviewees were drawing upon a wider range of experience, and these contrasting experiences and ways of sharing were of clear relevance to the research objectives. Looking at the findings, one drawback of this approach is that the discussion of governance processes is not grounded in specific cases that may have allowed more detailed exploration of the topic. I believe, however, that the trade-off of enabling a broad examination of the topic was better aligned with the overall research aims and objectives.

Finally, there are some groups of people that are less well represented in the study than they might have been. Not all of those invited to interview responded or agreed to be interviewed. This has resulted in some groups being less well represented than others – in particular policy makers in

LMICs are rather less well represented than they would have been ideally. In addition, at the time of data collection those from the commercial sectors were not noted to be prominent stakeholders, although they are increasingly so, and as a result were not included in the sampling frame. Although it appears that theoretical saturation was reached, there may be other groups who would have provided new perspectives that were not included in this study.

8.7. Future research

There are a number of avenues for future research raised by this study. Firstly, there are additional groups of people whose perspectives would be valuable for a holistic understanding of the topic. The most obvious of these is the patients and research participants whose data or samples may be shared during infectious disease responses. It is clear that interviewees saw this as a key group that needed protecting, and an important complement to this study would be one that ascertained their perspectives, across diverse geographical locations. A further group of people whose perspectives warrant further research, that were not included in this study as noted in the limitations, are those who work in industry, for example pharmaceutical and technology companies.

Secondly, given the dynamic nature of infectious diseases, it would be valuable to better understand how perspectives change over time. The data collection for this study was completed prior to the Covid-19 pandemic, and perspectives on certain issues may have changed as a result of experiences during this pandemic. There are also new questions that arise as a result of these situations, such as the ethical implications of data sharing in relation to contact tracing apps – which has been raised in some recent research (153, 154).

Thirdly, there are a number of novel issues that were raised but not resolved in this study that would benefit from further focused normative work. This includes:

- the ethical implications of the practical challenges to obtaining consent, engaging the community and completing robust ethics review in the face of an emergency;
- the ethical implications of where data servers are located;
- the distinct ethical issues relevant to sharing qualitative data during an infectious disease response;
- and unpacking individual and population interests when sharing data and samples in emergencies, with attentiveness to different types of data, and samples, and different types of disease.

Finally, there is scope for more research on good governance in relation to data and sample sharing respectively. This includes examination of the role and composition of access committees, and good practice in relation to the development of MTAs and DSAs.

8.8. Conclusions

Through this study I have shown that data and sample sharing for infectious disease research and responses raises some key ethical considerations for professionals and other stakeholders that both act to promote and constrain sharing. Crucially, this is a contested area where there are a range of perspectives on the nature of ethical practice. This study has highlighted several considerations for policy and practice, and future research.

The study data collection took place prior to the ongoing (at the time of writing) Covid-19 pandemic, which has shown how interconnected we are as a global community. The pandemic underscores the importance of examining ethical considerations in relation to the infectious disease response – a time when there is much uncertainty, and a great deal of pressure to act rapidly. Such pandemics are sadly expected to arise again, and it is key that global preparedness continues in

advance of the next event, including with respect to developing fair and inclusive systems for data and sample sharing.

My next step will be to continue to disseminate the findings through presentations and peer-reviewed publications.

References

1. World Health Organization. Urgent health challenges for the next decade. [30 September 2020]. Available from: https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade?utm_source=STAT+Newsletters&utm_campaign=1931cb646b-MR_COPY_02&utm_medium=email&utm_term=0_8cab1d7961-1931cb646b-150708293.

2. World Health Organization. Health topics. Infectious diseases. [13 September 2020]. Available from: https://www.who.int/topics/infectious_diseases/en/.

3. Roser M, Ritchie H. "Burden of Disease". Published online at OurWorldInData.org. 2016 [30 September 2020]. Available from: https://ourworldindata.org/burden-of-disease.

4. Institute for Health Metrics and Evaluation. Global Health Data Exchange. GBD Results Tool. [30 September 2020]. Available from: http://ghdx.healthdata.org/gbd-results-tool.

5. Diamond JM. Guns, germs, and steel : the fates of human societies: New York : Norton; 2005.

6. Crawford DH. Deadly companions : how microbes shaped our history. Oxford: Oxford University Press; 2007.

7. Davies SE, Kamradt-Scott A, Rushton S. Disease diplomacy: international norms and global health security. Baltimore, MD: Johns Hopkins University Press; 2015.

8. Bloom DE, Cadarette D. Infectious Disease Threats in the Twenty-First Century: Strengthening the Global Response. Frontiers in Immunology. 2019;10(549).

9. Yozwiak NL, Schaffner SF, Sabeti PC. Data sharing: Make outbreak research open access. Nature. 2015;518(7540):477.

10. McNabb SJ, Shaikh AT, Nuzzo JB, Zumla AI, Heymann DL. Triumphs, trials, and tribulations of the global response to MERS coronavirus. Lancet Respir Med. 2014;2(6):436-7.

11. U.S. Department of Energy. Human Genome Project Information Archive 1990–2003. Policies on Release of Human Genomic Sequence Data Bermuda-Quality Sequence [13 September 2020]. Available from: https://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml.

12. The Wellcome Trust. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. London: The Wellcome Trust; 2003.

13. Toronto International Data Release Workshop A. Prepublication data sharing. Nature. 2009;461(7261):168-70.

14. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials. BMJ. 2017;357:j2372.

15. Wellcome Trust. Sharing research data to improve public health: full joint statement by funders of health research. [2 September 2015]. Available from: http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm.

16. The Academy of Medical Sciences. Our data-driven future in healthcare. People and partnerships at the heart of health related technologies. London: The Academy of Medical Sciences; 2018.

17. Ascoli GA, Maraver P, Nanda S, Polavaram S, Armañanzas R. Win–win data sharing in neuroscience. Nature Methods. 2017;14(2):112-6.

18. Kaye J, Gibbons S, Heeney C, Smart A. Governing biobanks: understanding the interplay between law and practice: Bloomsbury Publishing; 2012.

19. Staunton C, de Vries J. The governance of genomic biobank research in Africa: reframing the regulatory tilt. Journal of Law and the Biosciences. 2020.

20. National Academies of Sciences Engineering Medicine. Sharing Research Data to Improve Public Health in Africa: A Workshop Summary. Connell ME, Plewes TJ, editors. Washington, DC: The National Academies Press; 2015.

21. Munung NS, Mayosi BM, de Vries J. Equity in international health research collaborations in Africa: Perceptions and expectations of African researchers. PLOS ONE. 2017;12(10):e0186237.

22. Bull S, Cheah PY, Denny S, Jao I, Marsh V, Merson L, et al. Best Practices for Ethical Sharing of Individual-Level Health Research Data From Low- and Middle-Income Settings. J Empir Res Hum Res Ethics. 2015;10(3):302-13.

23. Parker M, Bull SJ, de Vries J, Agbenyega T, Doumbo OK, Kwiatkowski DP. Ethical Data Release in Genome-Wide Association Studies in Developing Countries. PLOS Medicine. 2009;6(11):e1000143.

24. Sheehan M. Can Broad Consent be Informed Consent? Public Health Ethics. 2011;4(3):226-35.

25. Haga SB, Beskow LM. Ethical, Legal, and Social Implications of Biobanks for Genetics Research. Advances in Genetics. 60: Academic Press; 2008. p. 505-44.

26. Tindana P, Molyneux CS, Bull S, Parker M. Ethical issues in the export, storage and reuse of human biological samples in biomedical research: perspectives of key stakeholders in Ghana and Kenya. BMC Med Ethics. 2014;15:76.

27. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14:1144.

28. Bull S, Roberts N, Parker M. Views of Ethical Best Practices in Sharing Individual-Level Data From Medical and Public Health Research: A Systematic Scoping Review. J Empir Res Hum Res Ethics. 2015;10(3):225-38.

29. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International journal of social research methodology. 2005;8(1):19-32.

30. Armstrong R, Hall BJ, Doyle J, Waters E. 'Scoping the scope' of a cochrane review. Journal of Public Health. 2011;33(1):147-50.

31. World Health Organization. WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases. Geneva; 2000.

32. van Schalkwyk G, de Vries J, Moodley K. "It's for a good cause, isn't it?" - Exploring views of South African TB research participants on sample storage and re-use. BMC Medical Ethics Vol 13 Jul 2012, ArtID 19. 2012;13.

33. Wendler D, Pace C, Talisuna AO, Maiso F, Grady C, Emanuel EJ. Research on stored biological samples: The views of Ugandans. IRB. 2005;27(2):1-5.

34. Zhang X, Matsui K, Krohmal B, Zeid AA, Muthuswamy V, Koo YM, et al. Attitudes towards transfers of human tissue samples across borders: An international survey of researchers and policy makers in five countries. BMC medical ethics. 2010;11(1):1.

35. Patterson C, McDaid LM, Hilton S. Gay and bisexual men's perceptions of the donation and use of human biological samples for research: A qualitative study. PLoS ONE. 2015;10(6).

36. MacQueen KM, Alleman P. International Perspectives on the Collection, Storage, and Testing of Human Biospecimens in HIV Research. IRB. 2008;30(4):9-14.

37. Modjarrad K, Moorthy VS, Millett P, Gsell P-S, Roth C, Kieny M-P. Developing Global Norms for Sharing Data and Results during Public Health Emergencies. PLoS Med. 2016;13(1):e1001935.

38. Garcia-Merino I, Cuevas Ndl, Jimenez JL, Garcia A, Gallego J, Gomez C, et al. Pediatric HIV BioBank: a new role of the Spanish HIV BioBank in Pediatric HIV research. AIDS research and human retroviruses. 2010;26(2):241-4.

39. Garcia-Merino I, de Las Cuevas N, Jimenez JL, Gallego J, Gomez C, Prieto C, et al. The Spanish HIV BioBank: a model of cooperative HIV research. Retrovirology. 2009;6:27.

40. Merino IMG, Fernandez IC, Lomba I, Torre AG, De La Fuente JG, Rico CG, et al. HIV HGM Spanish biobank: Ethical and legal conflitcs in the yield of biobanks. Biopreservation and Biobanking. 2011;9 (3):305.

41. Munoz-Fernandez MA, Merino IG, Consuegra I, Jimenez JL. The Spanish HIV HGM BioBank (SHIVBB). Biopreservation and Biobanking. 2013;11(4):253-4.

42. Cason J, Kozlakidis Z. The Infectious Diseases BioBank (IDB) at King's College London, United Kingdom. Biopreservation and Biobanking. 2012;10(3):295-6.

43. Towie N. London hospital launches infectious disease 'biobank'. Nature Medicine. 2007;13(6):653.

44. Williams R, Mant C, Cason J. The Infectious Diseases BioBank at King's College London: archiving samples from patients infected with HIV to facilitate translational research. Retrovirology. 2009;6:98.

45. Fidler DP. Who owns MERS? The intellectual property controversy surrounding the latest pandemic. Foreign Affairs June. 2013;7.

46. Aldis W. Health security as a public health concept: a critical analysis. Health Policy & Planning. 2008;23(6):369-75.

47. Keuleyan E. Liberty to decide on dual use biomedical research: An acknowledged necessity. Science and engineering ethics. 2010;16(1):43-58.

48. Fairchild AL, Gable L, Gostin LO, Bayer R, Sweeney P, Janssen RS. Public Goods, Private Data: HIV and the History, Ethics, and Uses of Identifiable Public Health Information. Public Health Reports. 2007;122(Suppl 1):7-15.

49. Aarestrup FM, Koopmans MG. Sharing Data for Global Infectious Disease Surveillance and Outbreak Detection. Trends Microbiol. 2016;24(4):241-5.

50. Langat P, Pisartchik D, Silva D, Bernard C, Olsen K, Smith M, et al. Is there a duty to share? Ethics of sharing research data in the context of public health emergencies. Public Health Ethics. 2011;4(1):4-11.

51. Schroeder D, Gefenas E. Realizing Benefit Sharing - The Case of Post-Study Obligations. Bioethics. 2012;26(6):305-14.

52. Andanda PA. Human-tissue-related inventions: ownership and intellectual property rights in international collaborative research in developing countries. Journal of medical ethics. 2008;34(3):171-9.

53. Betsou F, Parida SK, Guillerm M. Infectious diseases biobanking as a catalyst towards personalized medicine: Mycobacterium tuberculosis paradigm. Tuberculosis. 2011;91(6):524-32.

54. Check Hayden E. Proposed Ebola biobank would strengthen African science. Nature. 2015;524(7564):146-7.

55. de Montgolfier S, Moutel G, Duchange N, Theodorou I, Hervé C, Leport C. Ethical reflections on pharmacogenetics and DNA banking in a cohort of HIV-infected patients. Pharmacogenetics. 2002;12(9):667-75.

56. Fidler DP. Influenza Virus Samples, International Law, and Global Health Diplomacy. Emerging Infectious Diseases. 2008;14(1):88-94.

57. Gasner MR, Fuld J, Drobnik A, Varma JK. Legal and policy barriers to sharing data between public health programs in New York City: a case study. American Journal of Public Health. 2014;104(6):993-7.

58. Kamau EM. Field-derived research materials in the African region: challenges and opportunities for the malaria researcher. (Special issue: Malaria research in Africa - multilateral initiative on malaria.). Acta Tropica. 2005;95(3):167-71.

59. Karunakara U. Data Sharing in a Humanitarian Organization: The Experience of Medecins Sans Frontires. PLoS Med. 2013;10(12):e1001562.

60. Kelley M. Infectious disease research and dual-use risk. Virtual Mentor. 2006;8(4):230-4.

61. McCloskey B, Dar O, Zumla A, Heymann DL. Emerging respiratory tract infections 2: emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. Lancet Infectious Diseases. 2014;14(10):1001-10.

62. Moorthy VS, Roth C, Olliaro P, Dye C, Kieny MP. Best practices for sharing information through data platforms: establishing the principles. Bulletin of the World Health Organization. 2016;94(4):234-A.

63. Sedyaningsih ER, Isfandari S, Soendoro T, Supari SF. Towards mutual trust, transparency and equity in virus sharing mechanism: the avian influenza case of Indonesia. Annals of the Academy of Medicine, Singapore. 2008;37(6):482-8.

64. The Plos Medicine Editors. Can Data Sharing Become the Path of Least Resistance? PLoS Med. 2016;13(1):e1001949.

65. Vallance P, Freeman A, Stewart M. Data Sharing as Part of the Normal Scientific Process: A View from the Pharmaceutical Industry. PLoS Med. 2016;13(1):e1001936.

66. Vogel G. Delays hinder Ebola genomics. Science. 2014;346(6210):684-5.

67. Vong S, Feng Z. Early response to the emergence of influenza A (H7N9) virus in humans in China: the central role of prompt information sharing and public communication. Bulletin of the World Health Organization. 2014;92(4):303-8.

68. Sharing saves lives. Nat Rev Micro. 2006;4(10):718-.

69. Editorial team Collective. World Health Organization to meet again on influenza virus-sharing. Euro Surveill. 2007;12(10):E071018.4.

70. Bogner P, Capua I, Lipman DJ, Cox NJ, others. A global initiative on sharing avian flu data. Nature. 2006;442(7106):981-.

71. Bond KC, Macfarlane SB, Burke C, Ungchusak K, Wibulpolprasert S. The Evolution and Expansion of Regional Disease Surveillance Networks and Their Role in Mitigating the Threat of Infectious Disease Outbreaks. Emerging Health Threats Journal. 2013;6:10.3402/ehtj.v6i0.19913.

72. Calain P, Fiore N, Poncin M, Hurst SA. Research Ethics and International Epidemic Response: The Case of Ebola and Marburg Hemorrhagic Fevers. Public Health Ethics. 2009;2(1):7-29.

73. Callaway E. Zika-microcephaly paper sparks data-sharing confusion. Nature News. 2016.

74. Caplan AL, Curry DR. Leveraging Genetic Resources or Moral Blackmail? Indonesia and Avian Flu Virus Sample Sharing. The American Journal of Bioethics. 2007;7(11):1-2.

75. Fidler DP, Gostin LO. The WHO pandemic influenza preparedness framework: a milestone in global governance for health. JAMA. 2011;306(2):200-1.

76. Heymann DL, Liu J, Lillywhite L. Partnerships, Not Parachutists, for Zika Research. New England Journal of Medicine. 2016;374(16):1504-5.

77. Phommasack B, Jiraphongsa C, Oo MK, Bond KC, Phaholyothin N, Suphanchaimat R, et al. Mekong Basin Disease Surveillance (MBDS): a trust-based network. (Special Issue: A world united against infectious diseases: Connecting Organizations for Regional Disease Surveillance (CORDS).). Emerging Health Threats Journal. 2013;6(19944).

78. Schopper D, Upshur R, Matthys F, Singh JA, Bandewar SS, Ahmad A, et al. Research Ethics Review in Humanitarian Contexts: The Experience of the Independent Ethics Review Board of Médecins Sans Frontières. PLoS Med. 2009;6(7):e1000115.

79. Sane J, Edelstein M. Overcoming Barriers to Data Sharing in Public Health A Global Perspective. London: The Royal Institute for International Affairs; 2015.

80. Merson L, Gaye O, Guerin PJ. Avoiding Data Dumpsters — Toward Equitable and Useful Data Sharing. New England Journal of Medicine. 2016;374:2414-5.

81. Ross E. Perspectives on Data Sharing in Disease Surveillance. London: The Royal Institute for Internaitonal Affairs; 2014.

82. Abayomi A, Christoffels A, Grewal R, Karam LA, Rossouw C, Staunton C, et al. Challenges of biobanking in South Africa to facilitate indigenous research in an environment burdened with human immunodeficiency virus, tuberculosis, and emerging noncommunicable diseases. Biopreservation and Biobanking. 2013;11(6):347-54.

83. Butler D. Tensions linger over discovery of coronavirus. Nat News. 2013.

84. Fidler DP. Negotiating Equitable Access to Influenza Vaccines: Global Health Diplomacy and the Controversies Surrounding Avian Influenza H5N1 and Pandemic Influenza H1N1. PLoS Medicine. 2010;7(5):e1000247.

85. Whitty CJM, Mundel T, Farrar J, Heymann DL, Davies SC, Walport MJ. Providing incentives to share data early in health emergencies: the role of journal editors. The Lancet. 2015;386(10006):1797-8.

86. Enserink M. Indonesia Earns Flu Accord at World Health Assembly. 2007;316(5828):1108-.

87. Brewer DD, Potterat JJ, Muth SQ. Withholding access to research data. Lancet. 2010;375(9729):1872; author reply 3.

88. World Health Organization. Ethics in epidemics, emergencies and disasters: Research, surveillance and patient care. WHO training manual. Geneva: World Health Organization; 2015.

89. Petro JB, Relman DA. Understanding Threats to Scientific Openness. Science. 2003;302(5652):1898-.

90. Gresham L, Ramlawi A, Briski J, Richardson M, Taylor T. Trust Across Borders: Responding to 2009 H1N1 Influenza in the Middle East. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science. 2009;7(4):399-404.

91. Kimball AM, Moore M, French HM, Arima Y, Ungchusak K, Wibulpolprasert S, et al. Regional infectious disease surveillance networks and their potential to facilitate the implementation of the international health regulations. Medical Clinics of North America. 2008;92(6):1459-71.

92. Médecins Sans Frontières. MSF Data sharing policy. Geneva: Médecins Sans Frontières; 2013.

93. World Health Organization. Frequently asked questions about the International Health Regulations (2005). 2009 [9 January 2017]. Available from: http://www.who.int/ihr/about/faq/en/.

94. World Health Organization. International Health Regulations (2005). Third ed. Geneva: World Health Organization; 2016.

95. Heymann DL, Chen L, Takemi K, Fidler DP, Tappero JW, Thomas MJ, et al. Global health security: the wider lessons from the west African Ebola virus disease epidemic. The Lancet. 2015;385(9980):1884-901.

96. World Health Organization. Health topics. Health security. [19 September 2020]. Available from: https://www.who.int/health-topics/health-security/#tab=tab_1.

97. United Nations Security Council resolution 1308, on the responsibility of the Security Council in the maintenance of international peace and security: HIV/AIDS and international peacekeeping operations, S/RES/1308 (17 July 2000).

98. Global Health Security Agenda. About. [9 January 2017]. Available from: https://www.ghsagenda.org/about.

99. Global Health Security Agenda. Membership. [9 January 2017]. Available from: https://www.ghsagenda.org/members.

100. Centers for Disease Control and Prevention. Global Health Security Agenda: Action Packages. 2014 [9 January 2017]. Available from: https://www.cdc.gov/globalhealth/healthprotection/ghs/pdf/ghsa-action-packages_24-september-2014.pdf.

101. Global Health Security Initiative. Welcome. [2 January 2017]. Available from: http://www.ghsi.ca/english/index.asp.

102. Calain P. Exploring the international arena of global public health surveillance. Health policy and planning. 2007;22(1):2-12.

103. Calain P. From the field side of the binoculars: a different view on global public health surveillance. Health policy and planning. 2007;22(1):13-20.

104. Convention on Biological Diversity. History of the Convention [9 January 2017]. Available from: https://www.cbd.int/history/.

105. United Nations. Convention on Biological Diversity. 1992.

106. Secretariat of the Convention on Biological Diversity. Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity : text and annex. Montreal: Secretariat of the Convention on Biological Diversity; 2011.

107. World Health Organization. Pandemic influenza preparedness Framework for the sharing of influenza viruses and access to vaccines and other benefits. Geneva: World Health Organization; 2011.

108. World Health Organization. Developing Global Norms for Sharing Data and Results during Public Health Emergencies. [10 January 2017]. Available from: http://www.who.int/medicines/ebola-treatment/data-sharing_phe/en/.

109. World Health Organization. Developing global norms for sharing data and results during public health emergencies. Statement arising from a WHO Consultation held on 1-2 September 2015. [10 January 2017]. Available from: http://www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/en/.

110. World Health Organization. Guidance for managing ethical issues in infectious disease outbreaks. Geneva: World Health Organization; 2016.

111. Pan American Health Organization. Zika Ethics Consultation: Ethics Guidance on Key Issues Raised by the Outbreak. Washington, DC: Pan American Health Organization; 2016.

112. The Wellcome Trust. Statement on data sharing in public health emergencies 2016 [10 January 2017]. Available from: https://wellcome.ac.uk/what-we-do/our-work/statement-data-sharing-public-health-emergencies.

113. International Severe Acute Respiratory and Emerging Infection Consortium. About Us. [10 January 2017]. Available from: https://isaric.tghn.org/about/.

114. Global Research Collaboration for Infectious Disease Preparedness. Governance. [10 January 2017]. Available from: http://www.glopid-r.org/learn-about-us/governance/.

115. International Severe Acute Respiratory and Emerging Infection Consortium. ISARIC Sample and Data Sharing Policy. Version 4. 2014 [29 September 2020]. Available from: https://isaric.tghn.org/site_media/media/articles/ISARIC_Sample_and_Data_Sharing_Policy_V4.pdf.

116. de Vries J, Williams TN, Bojang K, Kwiatkowski DP, Fitzpatrick R, Parker M. Knowing who to trust: exploring the role of 'ethical metadata' in mediating risk of harm in collaborative genomics research in Africa. BMC Med Ethics. 2014;15:62.

117. Parker M, Kingori P. Good and Bad Research Collaborations: Researchers' Views on Science and Ethics in Global Health Research. PLOS ONE. 2016;11(10):e0163579.

118. Rathi V, Dzara K, Gross CP, Hrynaszkiewicz I, Joffe S, Krumholz HM, et al. Sharing of clinical trial data among trialists: a cross sectional survey. BMJ : British Medical Journal. 2012;345:e7570.

119. Bryman A. Social research methods. 4th ed. Oxford: Oxford University Press; 2012.

120. Pisani E, Botchway S. Sharing individual patient and parasite-level data through the WorldWide Antimalarial Resistance Network platform: A qualitative case study [version 1; peer review: 2 approved]. Wellcome Open Research. 2017;2(63).

121. World Health Organization. Leishmaniasis. Clinical forms of the leishmaniases. [23 July 2020]. Available from: https://www.who.int/leishmaniasis/disease/clinical_forms_leishmaniases/en/index2.html.

122. World Health Organization. Leishmaniasis. Magnitude of the problem. [23 July 2020]. Available from: https://www.who.int/leishmaniasis/burden/magnitude/burden_magnitude/en/.

123. Infectious Diseases Data Observatory. Visceral Leishmaniasis [30 August 2020]. Available from: https://www.iddo.org/research-themes/visceral-leishmaniasis.

124. Infectious Diseases Data Observatory. Visceral Leishmaniasis. Strutural governance. [30 August 2020]. Available from: https://www.iddo.org/vl/governance/structural-governance.

125. World Health Organization. Ebola virus disease. [30 August 2020]. Available from: https://www.who.int/health-topics/ebola/#tab=tab_1.

126. Centers for Disease Control and Prevention. Ebola (Ebola Virus Disease). 2014-2016 Ebola Outbreak in West Africa. 2019 [30 August 2020]. Available from: https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html.

127. Infectious Diseases Data Observatory. Ebola. [30 August 2020]. Available from: https://www.iddo.org/research-themes/ebola.

128. Infectious Diseases Data Observatory. Ebola. Structure. [30 August 2020]. Available from: https://www.iddo.org/ebola/governance/structure.

129. Lo Iacono V, Symonds P, Brown DHK. Skype as a Tool for Qualitative Research Interviews. Sociological Research Online. 2016;21(2):103-17.

130. Sullivan JR. Skype: An Appropriate Method of Data Collection for Qualitative Interviews? The Hilltop Review. 2012;6(1):54-60.

131. Weller S. The potentials and pitfalls of using Skype for qualitative (longitudinal) interviews. NCRM Workng Paper. NCRM. 2015.

132. Le Quéré CC, Stuart; Corner, Adam; Cutting, David; Johnson, Martin; Minns, Asher; Schroeder, Heike; Walker-Springett, Kate; Whitmarsh, Lorraine; Wood, Ruth. Towards a culture of low-carbon research for the 21st Century. Tyndall Centre for Climate Change Research. Working Paper 161. 2015.

133. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.

134. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). Medical Education. 2006;40(5):405-14.

135. Bosk CL. What Would You Do?: Juggling Bioethics and Ethnography: University of Chicago Press; 2008.

136. Wong G, Liu W, Liu Y, Zhou B, Bi Y, Gao George F. MERS, SARS, and Ebola: The Role of Super-Spreaders in Infectious Disease. Cell Host & Microbe. 2015;18(4):398-401.

137. Hou Y, Zhao J, Martin W, Kallianpur A, Chung MK, Jehi L, et al. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. BMC Medicine. 2020;18(1):216.

138. U.S. Department of Health & Human Services. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [24 September 2020]. Available from: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.

139. World Health Organization. Global Observatory on Health R&D. WHO R&D Blueprint [5 September 2020]. Available from: https://www.who.int/research-observatory/analyses/rd_blueprint/en/.

140. World Health Organization. Immunization, Vaccines and Biologicals. Public health surveillance [6 September 2020]. Available from: https://www.who.int/immunization/monitoring_surveillance/burden/vpd/en/.

141. World Health Organization. What are the Emergency Medical Teams? [6 September 2020]. Available from: https://www.who.int/news-room/q-a-detail/emergency-medical-teams.

142. Global Research Collaboration for Infectious Disease Preparedness. About Us. [6 September 2020]. Available from: https://www.glopid-r.org/about-us/.

143. Nuffield Council on Bioethics. Research in global health emergencies: ethical issues. London: Nuffield Council on Bioethics; 2020.

144. Fairhead J, Leach M, Small M. Where techno-science meets poverty: Medical research and the economy of blood in The Gambia, West Africa. Social Science & Medicine. 2006;63(4):1109-20.

145. Staunton C, Tindana P, Hendricks M, Moodley K. Rules of engagement: perspectives on stakeholder engagement for genomic biobanking research in South Africa. BMC Medical Ethics. 2018;19(1):13.

146. Lucassen A, Parker M. Confidentiality and sharing genetic information with relatives. The Lancet. 2010;375(9725):1507-9.

147. Marsh V, Kombe F, Fitzpatrick R, Williams TN, Parker M, Molyneux S. Consulting communities on feedback of genetic findings in international health research: sharing sickle cell disease and carrier information in coastal Kenya. BMC medical ethics. 2013;14:41-.

148. Hall W, Carter L, Morley KI. Neuroscience research on the addictions: A prospectus for future ethical and policy analysis. Addictive Behaviors. 2004;29(7):1481-95.

149. Moodley K, Sibanda N, February K, Rossouw T. "It's my blood": ethical complexities in the use, storage and export of biological samples: perspectives from South African research participants. BMC Medical Ethics. 2014;15(1):4.

150. Buchanan EA, Zimmer M. "Internet Research Ethics", The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.) [29 September 2020]. Available from: https://plato.stanford.edu/archives/win2018/entries/ethics-internet-research/.

151. Mozersky J, Parsons M, Walsh H, Baldwin K, McIntosh T, DuBois JM. Research Participant Views regarding Qualitative Data Sharing. Ethics & Human Research. 2020;42(2):13-27.

152. DuBois JM, Strait M, Walsh H. Is it time to share qualitative research data? Qualitative Psychology. 2018;5(3):380-93.

153. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. Journal of Medical Ethics. 2020;46(7):427.

154. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. The Lancet Digital Health. 2020;2(8):e425-e34.

155. Gewin V. Six tips for data sharing in the age of the coronavirus. Nature. 2020.

156. Kerasidou A. The role of trust in global health research collaborations. Bioethics. 2019;33(4):495-501.

157. Dyer C. Covid-19: Rules on sharing confidential patient information are relaxed in England. BMJ. 2020;369:m1378.

158. Freudenthal E. Ebola's lost blood: row over samples flown out of Africa as 'big pharma' set to cash in. The Telegraph. 2019.

159. World Health Organization. Essential medicines and health products. Report on the 2nd WHO Consultation on Biobanking: Focus on West Africa. 2015 [5 Septeber 2020]. Available from: https://www.who.int/medicines/ebola-treatment/meetings/2nd_who_biobaking-consultation/en/.

160. UNAIDS. Criminalisation of HIV Non-Disclosure, Exposure and Transmission: Backgroung and Current Landscape. Geneva: UNAIDS; 2012.

161. ALERRT - African coaLition for Epidemic Research Response and Training. [5 September 2020]. Available from: https://www.alerrt.global/.

162. Médecins Sans Frontières. Pushed to the limit and beyond: a year into the largest ever Ebola outbreak. 2015.

163. Moon S, Sridhar D, Pate MA, Jha AK, Clinton C, Delaunay S, et al. Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola. The Lancet. 2015;386(10009):2204-21.

164. Infectious Diseases Data Observatory. Ebola. Our Work. [5 September 2020]. Available from: https://www.iddo.org/ebola/about-us/our-work.

165. Ipsos MORI. The One-Way Mirror: Public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. London: Ipsos MORI; 2016.

166. Global Research Collaboration for Infectious Disease Preparedness. Preparedness. Data Sharing. [20 August 2020]. Available from: https://www.glopid-r.org/our-work/data-sharing/.

167. Global Research Collaboration for Infectious Disease Preparedness. Data Sharing Working Group. [20 August 2020]. Available from: https://www.glopid-r.org/data-sharing-working-group-4/.

168. Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and health care: ethical issues. London: Nuffield Council on Bioethics; 2015.

169. Yassin R, Lockhart N, González del Riego M, Pitt K, Thomas JW, Weiss L, et al. Custodianship as an ethical framework for biospecimen-based research. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19(4):1012-5.

170. Fullerton SM, Anderson NR, Guzauskas G, Freeman D, Fryer-Edwards K. Meeting the Governance Challenges of Next-Generation Biorepository Research. Science Translational Medicine. 2010;2(15):15cm3.

171. Bhutta ZA. Beyond informed consent. Bull World Health Organ. 2004;82(10):771-7.

172. Participants in the Community Engagement Consent Workshop Kilifi Kenya March 2011. Consent and community engagement in diverse research contexts. Journal of empirical research on human research ethics : JERHRE. 2013;8(4):1-18.

173. Diallo DA, Doumbo OK, Plowe CV, Wellems TE, Emanuel EJ, Hurst SA. Community Permission for Medical Research in Developing Countries. Clinical Infectious Diseases. 2005;41(2):255-9.

174. Campbell MM, Susser E, de Vries J, Baldinger A, Sibeko G, Mndini MM, et al. Exploring researchers' experiences of working with a researcher-driven, population-specific community advisory board in a South African schizophrenia genomics study. BMC Medical Ethics. 2015;16(1):45.

175. Mlambo CK, Vernooij E, Geut R, Vrolings E, Shongwe B, Jiwan S, et al. Experiences from a community advisory Board in the Implementation of early access to ART for all in Eswatini: a qualitative study. BMC Medical Ethics. 2019;20(1):50.

176. Wright K, Sheather J. The Role of Solidarity in Research in Global Health Emergencies. The American Journal of Bioethics. 2020;20(5):4-6.

177. Zhao Y, Fitzpatrick T, Wan B, Day S, Mathews A, Tucker JD. Forming and implementing community advisory boards in low- and middle-income countries: a scoping review. BMC Medical Ethics. 2019;20(1):73.

178. Saxena A, Horby P, Amuasi J, Aagaard N, Köhler J, Gooshki ES, et al. Ethics preparedness: facilitating ethics review during outbreaks - recommendations from an expert panel. BMC Medical Ethics. 2019;20(1):29-.

179. Norton A, Pardinaz-Solis R, Carson G. GloPID-R Roadmap for Data Sharing in Public Health Emergencies. GLOPID-R.

180. Chatham House. Ethical Principles [5 September 2020]. Available from: https://datasharing.chathamhouse.org/ethical-principles/.

181. Star SL, Griesemer JR. Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science. 1989;19(3):387-420.

182. Star SL. This is Not a Boundary Object: Reflections on the Origin of a Concept. Science, Technology, & Human Values. 2010;35(5):601-17.

183. Bister MD. Human Tissue Research: Oxford University Press; 2011.

184. Kvale L, Pharo N. Understanding the Data Management Plan as a Boundary Object through a Multi-stakeholder perspective. International Journal of Digital Curation. 2020;15(1).

185. World Health Organization. Draft R&D Blueprint MTA tool. [5 September 2020]. Available from: https://apps.who.int/blueprint/mta-tool/index.html.

186. Carlile PR. A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. Organization Science. 2002;13(4):442-55.

Appendices

Appendix 1: Key words and MeSH terms

Population (optional)	'Exposure'	Outcome (phenomena of interest)
KEYWORDS	KEYWORDS	KEYWORDS
Developing countries	Public Health Emergency	Sharing Or release Or dissemination Or distribution Or export
Africa	Disaster	AND
Asia	Infectious disease	Data Or information Or results Or findings Or surveillance
Latin America	Communicable disease	OR
	Outbreak	Tissue Or Biological Samples Or Blood Or DNA Or Biological Materials Or Specimen Or Biobank
	Epidemic	
	Pandemic	
	Ebola	
	Influenza or flu	
	Marburg	
	Dengue	
	Human immunodeficiency virus or HIV or HIV/AIDS	
	Poliomyelitis or polio	
	Cholera	
	Yellow fever	
	Plague	
	Meningococcal or Neisseria meningitidis or meningococc*mia	
	middle east respiratory virus or middle east respiratory syndrome	
	Tuberculosis or TB	
	Malaria	
	Leishmaniasis or leishmaniosis or leishmania	
	Trypanosomiasis or trypanosomosis or trypanosoma	
	H*morrhagic fever	
MESH	MESH	MESH
Developing countries	Public Health	Information Dissemination
World Health	Emergencies/epidemiology*	Access to Information
Africa	Disaster	Data Collection
Asia	Communicable Diseases	Data Interpretation
	Disease Outbreaks	Data Mining

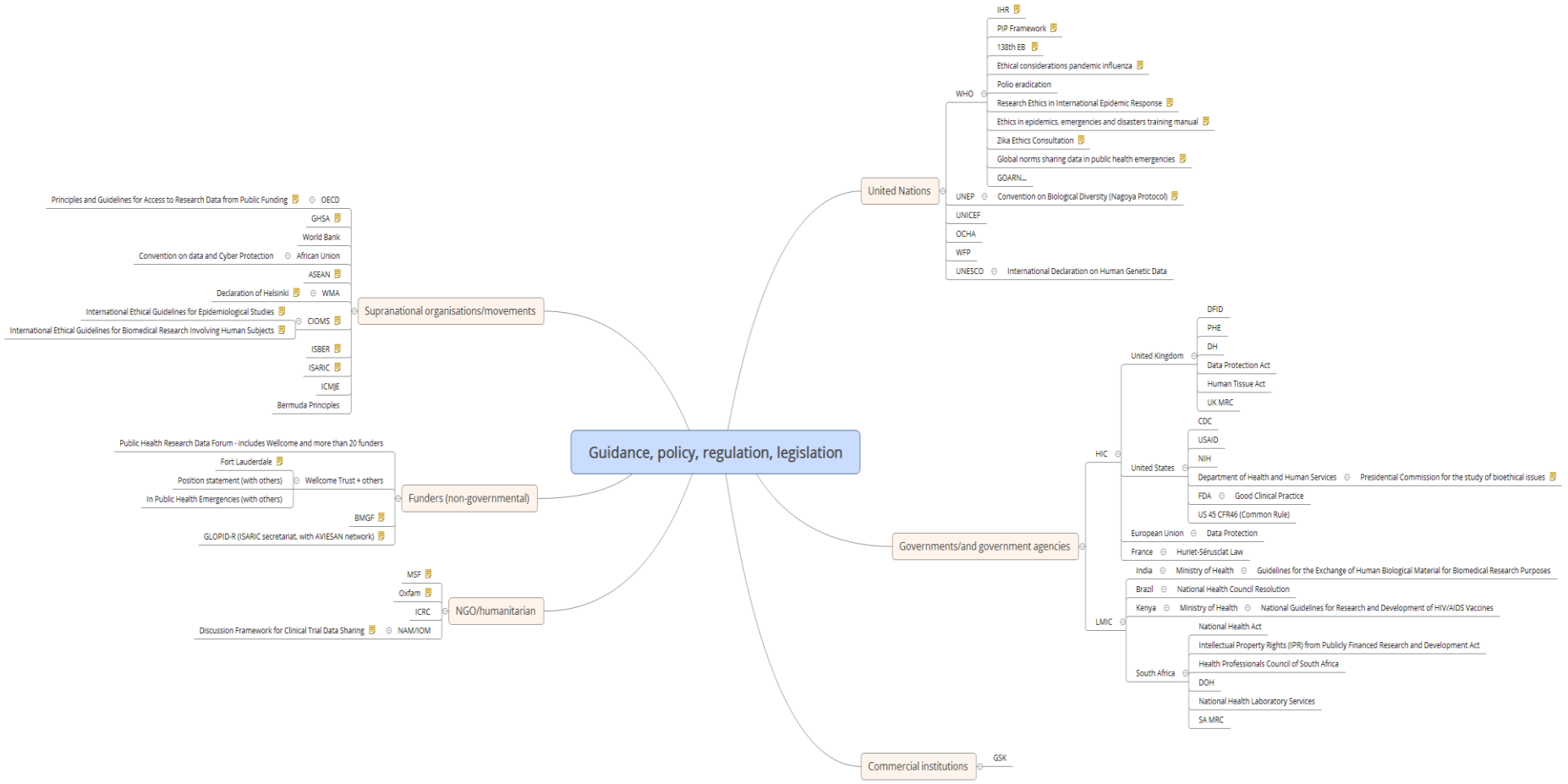
	Epidemics	Follow-Up Studies
	Pandemics	Communication
	Influenza, Human	Biological Products
	Hemorrhagic Fever, Ebola	Tissues
	Hemorrhagic Fevers, Viral	Blood
	Coronavirus/ and Respiratory Tract Infections	DNA
	Coronavirus Infections/	Blood Specimen Collection
	Marburg Virus Disease	Biological Specimen Banks
	Severe Dengue	
	Hemorrhagic Fever Virus, Crimean-Congo	
	Yellow Fever	
	Plague	
	Cholera	
	Meningococcal Infections	
	Meningitis, Meningococcal	
	Trypanosomiasis, African	
	HIV	
	Leishmaniasis	
	Poliomyelitis	
	Tuberculosis	
	Malaria	

Appendix 2: Medline search strategy

1. disasters/ or disaster planning/
2. Emergencies/ep [Epidemiology]
3. *Emergencies/
4. disease outbreaks/ or epidemics/ or pandemics/
5. Disease Transmission, Infectious/
6. exp Communicable Diseases/
7. Hemorrhagic Fever, Ebola/
8. Influenza, Human/
9. Coronavirus/ and Respiratory Tract Infections/
10. Coronavirus Infections/
11. Dengue/ or Severe Dengue/ or Dengue Virus/
12. Trypanosomiasis/
13. Poliomyelitis/ or Poliovirus/
14. Leishmaniasis/
15. HIV/ or HIV Infections/
16. Meningococcal Infections/ or Meningitis, Meningococcal/
17. Hemorrhagic Fevers, Viral/
18. Marburg Virus Disease/
19. Hemorrhagic Fever Virus, Crimean-Congo/
20. Yellow Fever/
21. Plague/
22. Cholera/
23. Tuberculosis/
24. Malaria/
25. ((disease* or public health or infect*) adj2 emergenc*).ti,ab.
26. ((disease or public health or infect*) adj2 outbreak?).ti,ab.
27. disaster?.ti,ab.
28. outbreak?.ti.
29. (epidemic? or pandemic?).ti,ab.
30. ((infectious or communicable) adj disease?).ti,ab.
31. ebola.ti,ab.
32. (influenza or flu).ti,ab.
33. (middle east respirat* adj (virus or syndrome?)).ti,ab.
34. marburg.ti,ab.
35. dengue.ti,ab.
36. polio*.ti,ab.
37. (human immunodeficiency virus or HIV*).ti,ab.
38. cholera.ti,ab.
39. yellow fever.ti,ab.
40. plague.ti,ab.
41. (meningococc* or neisseria meningitidis or meningitis).ti,ab.
42. (tuberculosis or TB).ti,ab.
43. malaria.ti,ab.
44. leishman*.ti,ab.
45. Trypanosom*.ti,ab.
46. h*morrhagic fever.ti,ab.

47. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46
48. specimen handling/ or biopsy/ or blood specimen collection/
49. Tissues/
50. Blood/
51. DNA/
52. ((tissue or blood or biolog* or dna) adj3 (sample? or specimen? or material?)).ti,ab.
53. ((public health or surveillance or clinical or routine or research) adj data).ti,ab.
54. 48 or 49 or 50 or 51 or 52 or 53
55. Information Dissemination/
56. Access to Information/
57. Data Collection/
58. "information storage and retrieval"/ or data mining/
59. Data Interpretation, Statistical/
60. confidentiality/ or genetic privacy/ or informed consent/
61. (sharing or governance or accountability or reciprocity or capacity building or partnership? or cooperat* or co-operat* or consent or confidentiality or cultural issue? or cultural sensitivit*).ti,ab.
62. 55 or 56 or 57 or 58 or 59 or 60 or 61
63. 54 and 62
64. Biological Specimen Banks/ or Tissue Banks/
65. (biobank* or bio bank* or biological bank* or tissue bank* or specimen bank*).ti,ab.
66. 63 or 64 or 65
67. 47 and 66

Appendix 3: Mind map of policy documents



Appendix 4: Participant information sheet

Study title: Sharing infectious disease related data and biological samples from low resource settings: perspectives on ethics and governance

This study is part of a doctoral research project being undertaken by Ruchi Baxi at the University of Oxford, about initiatives for sharing infectious disease related data and biological samples.

Purpose of this study

This study will use interviews to understand what people do and think in relation to their work on data or sample sharing, the types of issues that come up in their work, and how processes are developed to manage these issues. The findings from this study will be used to identify considerations for ethics frameworks on this topic.

This information sheet will help you to decide whether you want to be interviewed as a part of this study. You are free to say no without explanations. It will not be reported to anybody in your institution or have any influence on your present or future involvement with your institution. If you agree to take part, then you will be asked to sign a consent form.

Your part in the project

- You have been invited to take part in this study because you work on an area related to sharing public health related data or samples, or have done so in the past
- If you agree to participate you will be interviewed by Ruchi Baxi for about 60 to 90 minutes. This will either be face-to face in private at a convenient location, or using video calling facilities. The discussion will be audio recorded, if you agree.
- It is important to hear your personal views for this research. However, if you do not feel like talking about certain issues, you are free not to do so. The interview can be terminated at any time.
- You can withdraw from the study at any time. If you withdraw prior to analysis of the interview data, the information from your interview will not be included in the study. Once analysis has been undertaken it is very difficult to separate interview data from individual participants.

Possible risks and benefits

Talking about this topic can raise issues that may be sensitive. The aim of the study is to understand experiences and issues in relation to data and sample sharing, and not to make judgements about your work. Everything you say will remain confidential, and all recordings and reports will be kept safely. Your name and other people's names will not be shared with anybody else, and they will not be included in study reports. Quotes from your interview will be anonymised and you will have the opportunity to ensure they have been represented in a respectful manner. Your views and observations will not be available to any of your colleagues or managers. If you would like to, you will have the opportunity to alter statements you have made.

I hope that your participation can be useful to yourself, because it will help understand your point of view on data and/or sample sharing, and through this contribute to research on the topic.

If you have any questions about the study, please ask me at any time.

Contact Information for Questions or Concerns

The study has been approved by the University of Oxford OxTREC Ethics Review committee.

If you have any questions or concerns about any aspect of this study, please feel free to contact:

Dr Ruchi Baxi
University of Oxford
The Ethox Centre
Nuffield Department of Population Health
Old Road Campus
Oxford
OX3 7LF
ruchi.baxi@ethox.ox.ac.uk
http://www.ndph.ox.ac.uk/team/ruchi-baxi

Research supervisors:

Prof Michael Parker: michael.parker@ethox.ox.ac.uk
Dr Patricia Kingori: patricia.kingori@ethox.ox.ac.uk
Dr Premila Webster: premila.webster@dph.ox.ac.uk

You may also contact the University of Oxford Clinical Trials and Research Governance (CTRG) email: ctrg@admin.ox.ac.uk

Appendix 5: Consent Form

Study title: Sharing infectious disease related data and biological samples from low resource settings: perspectives on ethics and governance

Participant name:

I confirm that I have read and understood the participant information sheet. I understand that my participation in this study is voluntary. I also understand that I am free to withdraw from the study at any time, and that my withdrawal will not cause offense and it will not affect me in any way.

I agree to take part in the above named study.

I agree to audio recording of the interview Yes □ No □

Signed:

Date:

Contact Information for Questions or Concerns

If you have any questions or concerns about any aspect of this study, please feel free to contact:

Dr Ruchi Baxi
University of Oxford
The Ethox Centre
Nuffield Department of Population Health
Old Road Campus
Oxford
OX3 7LF
ruchi.baxi@ethox.ox.ac.uk
http://www.ndph.ox.ac.uk/team/ruchi-baxi

Research supervisors:

Prof Michael Parker: michael.parker@ethox.ox.ac.uk
Dr Patricia Kingori: patricia.kingori@ethox.ox.ac.uk
Dr Premila Webster: premila.webster@dph.ox.ac.uk

You may also contact the University of Oxford Clinical Trials and Research Governance (CTRG) email: ctrg@admin.ox.ac.uk

Appendix 6: Interview guide

Study title: Sharing infectious disease related data and biological samples from low resource settings: perspectives on ethics and governance

Interview details

Interview date: Audio recorded (Y/N):

Time at start: Time at end:

Place of interview:

Participant Information

1. Participant identifier:

2. Gender: 3. Age:

4. Nationality: 5. Country of residence:

6. Employer: 7. Role:

8: Involved with a particular initiative (Y/N), and which one(s):

Question areas[25]

General experience

1. What would you describe as your main role at work?
2. Have you been involved in much work on data and/or sample sharing? In what capacity?
3. How long ago did you get involved?
4. What led you to get involved?
5. How does this fit into your wider work?

Ideas around sharing

6. When you talk about data/sample sharing, what sorts of activities does this include?
   Probe further about activities and reasons, and ideas around the use of word 'sharing' in contrast to other terms
7. Do you feel sharing is important, and why?
8. In your view what are the main reasons for and against sharing data/samples?
9. What are/would your expectations be when contributing the data/samples?
10. What are/would your expectations be when receiving the data/samples?
11. Has your view of data/sample sharing changed over time?
12. Have you been involved in much sharing internationally? Have you been involved in much sharing in the same countries/nearby countries? How would you describe this/these experience(s)?

[25] Interview questions will be iteratively developed as the study progresses, incorporating arising issues and themes

P*robe about to what extent these are considered similar or distinct activities, and why*

13. Do you think that sharing data is a similar activity to sharing biological samples?
14. What do you see as 'successful' data/sample sharing? How would you describe 'failed' data/sample sharing?

Drivers/barriers

15. In your experience, how are initiatives started?
16. To your knowledge, who does most data/sample sharing occur between?
17. How do people become involved in initiatives?
18. How does the number of researcher groups in a given disease area shape data sharing in that disease area?
19. What do you see as the main factors that drive data/sample sharing?
20. What are the main barriers to data/sample sharing?
21. Do you think these barriers that should be and can be overcome? What do you think is needed to achieve this?

Ethical issues

22. What do you think are the main ethical issues in data/sample sharing?
    Probe issues raised further for reasons
23. Have you encountered many of these issues in practice? How have they manifested?
24. Are the issues that you have come across the sorts of things you were expecting when you started your work? Any surprises?
25. Do you think the issues vary from initiative to initiative? In your view, what are the factors behind this?
    If not raised, probe further on how issues relate to type of data/samples, nature of the infectious disease, other contextual factors such as origin (place, types of patients/participants), if part of an outbreak/emergency, likely users and uses of data/samples

Development of processes

26. How are requests for use of the data/samples managed?
27. How are ethical issues raised and discussed in initiatives you have been involved in?
28. How would you describe your role in managing the ethical issues?
29. What are the main types of processes used in the initiative(s) to manage the ethical issues?
    If not raised, ask about the role of policy, guidance, regulation, legislation
30. To your knowledge how are these processes developed? What are these processes based on?
31. What are the strengths and weaknesses of the processes?
32. In your view, are there other things that could be put in place to manage the ethical issues?
33. What are the main enablers to developing ethics processes?
34. What are the main barriers to developing ethics processes?

Appendix 7: List of codes

The codes that emerged through the analysis of interview data are listed below in alphabetical order, with their associated description. I exported the list from NVivo 12.

Name	Description
Academic rewards and recognition	Discussion of career advancement, publications, or royalties in academia/for researchers
Accountability	Discussion about role of accountability, and issues with respect to developing lines of accountability
Bad practice	Practice that is of concern because it is considered unethical, it violates norms or other guidance, or is otherwise poorly perceived
Barrier sharing data	Reasons that impede data/sample sharing, whether or not these are considered appropriate
Best practice guidance	Sources of guidance mentioned as a source of guidance, and any issues with these
Capacity building	Mention/discussion of the role of capacity building, good practice, and concerns
Changes over time	Reference to own views or general norms changing over time
Commercial and industry	Mention of pharmaceuticals, commercialisation or 'industry'
Confidentiality and anonymisation	Mention of these specifically for data/samples. Broader mention of 'privacy' will be coded separately.
Consent	Mention of consent or models of consent, and issues obtaining consent
Culture	Reference to different social or cultural perspectives on data/samples
Data access committee	Mention of the role of DACs, their utility, and any issues or concerns
Data contributors	Role in decisions about how data/samples are used
Data protection	Discussions around policy or practice, and how it relates to sharing

Data users	Mention of groups who use data (except commercial coded separately)
Data vs samples	Contrast between issues - both similarities and differences
Dealing with disagreement	Discussion of how conflict is managed e.g. involvement of third party
Definition of data sharing	How data/sample sharing is defined or described
Destruction of samples or data	Where discussed, that need to be destroyed, removed. Rationale for this.
Different infectious diseases	Distinctions or contrasts made between different infectious diseases
Dignity	Mention of dignity in relation to sharing
Distribution of benefits	Parent code
Academic benefits	Distinct from academic recognition and rewards (although related) as this is about fairness of the benefits
Benefits to communities	Benefits to patients/participants/general public
Duty of care	Reference to caring for the people that data or samples collected from
Ethics of not sharing	Mention of not sharing or withholding data/samples as an ethical issue
Exploitation	Discussion of exploitation specifically – often in tandem with the distribution of benefits. Code to be used where this is described as exploitative
Funders and funding	Mention of role of funders or issues related to funding, and positive or negative perspectives on this
Generalisability	Discussion of this topic, types of data and how they relate to this
Good practice	Discussion of practice that is considered appropriate, in line with accepted norms, and/or ethically good conduct
Harmonisation	Discussion around importance of data harmonisation, or issues with data that are not harmonised. This includes discussion of standardisation.

HIC and LMIC	Discussion around differences in considerations between HIC and LMICs, and any tensions where sharing occurs between HICs and LMICs
Hiding behind ethics	Discussion or suggestion that ethics is a tool to avoid sharing
History	Historical references to related to current attitudes to data/samples sharing, e.g. colonialism
Human rights	Specific mention of human rights, and how they relate to data/sample sharing
IHR	Mention of IHR, its use, and limitations
Incentivising data sharing	Actions taken to encourage or incentivise sharing
Incidental findings	Discussion of the issue of incidental findings, how to manage incidental findings that arise once data or samples are shared, and any issues or concerns in relation to this
Individuals vs public health	Discussion of where there are tensions or complementary aims to meet individual and public health interests, and how to manage these any tensions
Infectious disease context	Discussion about how infectious diseases as a group impact (or not) on data and sample sharing practice, particularly in comparison to sharing for other types of disease e.g. relevance of emergencies, or other features characteristic of infectious diseases
Intellectual property	Discussion of IP specifically, rather than commercialisation in broad terms
International sharing	Discussion of sharing outside of country in which data collected, and positive or negative perspectives on this
Journals	Role of journals in relation to data or sample sharing, and positive or negative perspectives on this
Keeping in country	Reference to keeping data/samples in the country of origin and rationale for this
Leadership	Discussion of the role of leadership, which organisations, groups or types of individuals are leaders, and any positive or negative perspectives on the leadership
Legal considerations	Reference to local, national or international law, or development of legal instruments

Media	Reference to the media, journalists in relation to infectious disease responses, and perspectives on how this relates to data or sample sharing
Misuse	Mention/explanation of term misuse or abuse, and related concerns
Moral character	Reference to the role and relevance of character in relation to ethical practice
Neutral decision making	Reference to neutrality or impartiality in decision-making as a part of the governance of data or sample sharing
Open data	Discussion on open data, or open data sharing
Ownership and control	Reference to owning data or samples, or control of them, including discussion of the rationale, and positive or negative perspectives on the role of ownership and control
Platform convener	Discussion around who should host data sharing platforms or biobanks
Politics	Mention of politics, how it impacts on data or sample sharing e.g. hesitation to share due to reputational concerns for the country
Power	Specific mention of power (not HIC/LMIC dynamic) e.g. hierarchy within teams impacting on decision-making
Prioritising sample use	Discussion on how to prioritise use of finite samples
Priority of data sharing	Relative importance given to data/sample sharing compared to other actions taken as part of the infectious disease response
Process - MTA or DSA	Reference to MTA or data transfer or data sharing agreement, and positive or negative perspectives on their use
Public engagement	Discussion about the role of public or community engagement
Public good	Mention of public good/interest in data or sample sharing, including how this is relevant to whether sharing ought to take place or not, and how it ought to take place
Qualitative vs quantitative	Discussion contrasting the issues sharing qualitative and quantitative data
Reasons for not sharing data	Reasons people give for withholding sharing data or samples

Reasons why people share	Why people may share data or samples that they have collected
RECs	Mention of role of RECs/IRBs/similar, and positive or negative experiences/perspectives
Repositories vs collaboration	Mention of contrast between sharing to a repository and collaboration
Research vs routine	Contrasts drawn between sharing for research and routine purposes, or the sharing of data or samples collected through routine or research activities
Respect for persons	Specific mention as a principle, and how it relates to sharing
Restrictions and controls	Different to ownership and control - about putting restrictions on access for certain reasons
Risks	Mention of term risk in general sense, and concerns that data or sample sharing is 'risky' and rationale for this
Sensitive data	Mention of the term 'sensitive' in relation to a type of data or sample, and any rationale given
Sharing with a team	Mention of sharing within a team as data or sample sharing
Speed	Discussion of the importance of speed, or concerns with time pressures
Stigma	Discussion of stigma, and how it relates to data/sample sharing
Terms of submission	Reference to the role of developing a terms of submission for practice
Trust	Mention of the role of relationships, trust or trustworthiness as a part of data or sample sharing, or of fear sharing with certain people (i.e. sign of mistrust)
Understanding ethics	Interviewees own self-perception about whether they understand what is encompassed in 'ethics'
Using own samples	Donation of own samples in a study
Value of data sharing	Parent code
Building capacity	Different to capacity building code - this is about discussing the value of data sharing as a prompt/catalyst for capacity building, rather than capacity building as a way to address imbalances

Clinical decision making	Not just general reference to 'improve patient care' but more specifically to influence a particular clinical decision
Enabling researchers	Enables more people to do research e.g. a source of research projects
Harness skills	Sharing data or samples as others may have particular skills
Influencing policy	Outputs of data or sample sharing and how they can influence or change policy
More or better science	Sharing to improve scientific knowledge/do 'better science'
Standardisation	Discussion of data sharing as a means to encourage harmonisation/standardisation of data collection
Optimising participant contribution	Discussion around making the most of data/samples participants have given their time/effort to contribute
Population or public health reasons	Discussion of benefits to either resulting from data/sample sharing
Sharing for quality assurance	Discussion of sharing to allow checking of analyses
Sharing to increase sample size	Discussion of sharing to improve sample sizes/potential to generate statistically significant findings, or do sub-group analyses
Transparency	Data/sample sharing as a way of promoting transparency in science/public health (similar to QA, put a more specific emphasis on the role of this in transparency)
Use of funds	Reference to making saving money/using funds effectively through re-using data or samples collected
Vulnerability	Mention of vulnerability of patients or research participants, and reasons for this, and implications of this