Introduction

> Water, water everywhere,
> Nor any drop to drink.[1]

THE PROBLEM AND THE POTENTIAL

The genesis of this project was the experience of the authors working with researchers, data custodians, ethics committees, and linkage units over the past decade or more managing access to government administrative data for research. Our individual backgrounds, experience, and expertise meant that we brought a diverse, interdisciplinary approach to the table. We started to identify the problems: we heard from researchers that there were concerns with unjustified delay and lack of access to data for research that was in the public interest; we heard from data custodians that they found the decision-making process difficult because they did not have sufficient resources or guidance; we heard from ethics committees that they found data linkage projects complex and difficult to assess; and we heard from consumer representatives that there were concerns about privacy, transparency, community awareness, and consultation.

We all understood the potential of this field of data research. We had experience working with passionate researchers who underlined for us the importance of using government administrative data for the greater good. We also understood the risks, having decades of expertise in privacy law and ethical review. Was it possible, we wondered, to achieve these great benefits while minimising the risks? What were the roadblocks, and could we help to clear the way? We all understood, to a greater or lesser degree, the mechanics of data linkage, having worked with and for the Australian Population Health Research Network (PHRN) in different capacities over the years as staff members, consultants, and members of advisory groups. We have watched at close hand this field of endeavour develop with ever increasing speed. The need for a better, more consistent approach to the regulatory frameworks

---

[1] S Coleridge, *The Rime of the Ancient Mariner* (Chatto & Windus, 1971).

and decision-making processes became clearer and more urgent as the problems of cross-jurisdictional linkage at the national level looked to become insurmountable at the international level unless a more consistent and sustainable model was developed.

While this book has strong theoretical foundations, we have focused on developing and presenting practical solutions to the problems faced by researchers, data custodians, ethics committees, the community, and other stakeholders in this complex and evolving environment. We have aimed to ensure that the discussion of specialist areas such as human rights, ethics, and law is accessible to everyone, rather than just experts in these fields. The future of public health research, and a great deal of other research essential to the well-being of the community, will rely on the use and linkage of data collections in the hands of government and beyond. This book aims to ensure that the field of research using linked data develops on sound ethical, rights-based, and lawful foundations. This will help to attract and sustain the social licence needed to support the use of this valuable community resource for the benefit of the community, locally, nationally, and globally.

Modern advances in information technology mean that we now live in a world awash with data collected from personal devices, public buildings and facilities, service providers, and other sources. In the last two decades, there has been a revolution in computational and statistical techniques so that it is now possible to analyse and make sense of these large volumes of data. The amount of data collected and the power to analyse and make use of it are set to increase exponentially. As a result, there are legitimate concerns about Big Data and the widespread collection and use of personal information by governments and the private sector.

A great deal of information about people is collected by governments, with and without their consent, while administering health programmes and providing public health services. While the data discussed is primarily health data, health research can be enhanced by linking health data to other personal information – such as education, justice, housing, and social welfare data – also held by government. All this information is sensitive and, while it is used in legitimate ways to administer and improve the public health system, it can also be used in illegitimate ways to profile and unjustly discriminate against individuals and groups. For this reason, the collection, use, and release of the information – including release for research – is highly regulated.

Despite this, there remains a danger that concerns about Big Data are having a detrimental impact on the beneficial use of government data collections for health and other research, conducted in highly regulated and privacy-protective environments. Access to population-wide collections of individual-level data is essential to support health research and to promote evidence-based health policy development. Use of existing data resources provides a time and cost-effective way to understand, monitor, and improve health and welfare. The challenge is to realise the potential benefits to health and health systems across the world offered by the use of this data, while respecting and responding to community concerns and minimising the risks to individuals and groups.

The focus of this book is on decision making. Access to government administrative data for research is based on a series of decisions by data custodians, ethics committees, and others that seek to balance the risks and benefits associated with the use of this data for research. These decisions operate as keys that unlock access for researchers to a well-guarded and extremely valuable community resource. It is critical that each key works well and is fit for purpose.

DATA LINKAGE

While many of the issues discussed in this book are relevant to all research using individual-level data we have chosen to focus on linked data. This is because linking data from different sources is the future of population health research. Linkage of population-wide data collections provides research resources that are inclusive, representative, accurate, and cost-effective as well as being an effective and efficient use of existing data.[2] It provides researchers with access to larger volumes of detailed data on individuals and groups, allowing for increasingly varied and comprehensive population-wide research.

Linkage of health data from different sources is occurring in a number of countries around the world, including across jurisdictional boundaries in federal and devolved systems such as Australia, Canada, and the United Kingdom (UK). This adds a layer of complexity to the ethical, legal, and policy arrangements. Working through this complexity is necessary to achieve nationally significant research results. Cross-jurisdictional linkage at the international level provides an even greater challenge. In a world of increasing population mobility, global epidemics of infectious diseases, and the shared responsibility for providing health care for everyone, it is essential to confront this challenge.

The COVID-19 pandemic provides the perfect example of our shared vulnerabilities and responsibilities and the need for a coordinated global response, including the linking of data across national boundaries. Linked data can be used to track the movement of the virus around the world, to learn about the progression of the disease and its symptoms, to evaluate the safety and efficacy of vaccines, to evaluate the efficacy of treatment, and to identify those who are most vulnerable for medical, socio-demographic, or geographical reasons.[3] An international understanding and approach to the ethical, legal, and policy issues will be required to support

---

[2] F Stanley, ‘Record Linkage: Public Good or Invasion of Privacy’ (Conference Paper, International Conference of Data Protection and Privacy Commissioners, 10 September 2003).

[3] PHRN, ‘Linked Data in the Fight against COVID-19’, *Impact Stories* (Web Page) <www.phrn.org.au/media/82041/impact-story-linked-data-in-the-fight-against-covid-19.pdf>.

Cambridge University Press & Assessment
978-1-108-42664-0 — Sharing Linked Data for Health Research
Carolyn Adams , Judy Allen , Felicity Flack
Excerpt
More Information

linkage across international borders so that the benefits of data linkage research at the national level can be translated to the global context. We hope that the principles and practice discussed in this book will provide a foundation for successful national and international linkage initiatives.

With the increased benefits of linked data, however, come increased risks to privacy, confidentiality, and security of the data. As the data collections to be linked increase in number, the legal requirements become more complex, as different legislation may apply to each data source linked. To address these increased risks and legal complexity, many countries have established specialist data linkage infrastructure and processes. These regional and national infrastructures take varied approaches to the range of ethical, legal, and policy issues inherent in data linkage research but there has been limited consideration of the strengths and weaknesses of the different approaches. One of the aims of this book is to examine these different approaches and consider what can be learned from them.

THE ETHICAL, LEGAL, AND PROCEDURAL ISSUES

The ethical, legal, and procedural concerns inherent in research using linked data include the impact of such research on individual autonomy and privacy and on relationships of confidentiality; the impact of research outcomes on individuals and collectives; the importance of data security; the difficulties with inconsistent and complex regulation across jurisdictions; requirements for approval from multiple ethics committees and data custodians; and the timeliness and transparency of decision making by data custodians and other stakeholders.

Some of these concerns stem from the fact that research using linked data involves whole populations and, because of this, individuals are not usually given the opportunity to consent to participation in the research. Instead, decisions are typically made on behalf of the community at the political level to establish data linkage infrastructure. The decision to permit access to linked data about individuals is then delegated to a variety of decision makers, including data custodians, ethics committees, and privacy committees. These regulatory systems have been developed over time to protect individual and organisational interests, but they are often complex and time-consuming. This has led to frustration among researchers and the decision makers themselves.

The aim of this book is to propose a framework for decision making about the linkage of individual-level data for research that will unlock critical public health benefits in a timely way while acknowledging and protecting the fundamental individual rights and community interests involved. The suggested framework of principles and practice is also intended to help maintain community support, or ‘social licence’, for the relevant decision making as well as for the resulting research.

SOCIAL LICENCE

The continued use of individual-level data without consent for research relies to a large extent on ongoing approval and broad social acceptance by the community and other stakeholders. The existing regulatory frameworks are sophisticated and complex; however, compliance with formal regulatory frameworks is not sufficient to establish a social licence for research unless the frameworks themselves adequately support social licence by ensuring trust and legitimacy.

The collection, use, and disclosure of individual information without individual consent – and generally without the knowledge of the individual – go beyond general norms of confidentiality and privacy. This is why community acceptance and support are critical for research using linked data. It needs to underpin the decisions made on behalf of the community that allow each step in the process, both in the public sector and in academic institutions. The role of private commercial interests either directly or as funders of research further complicates the picture.

Although social licence is intangible, without it research 'may experience ongoing challenge and contestation'.[4] This book examines the current regulatory systems in light of this claim and considers whether they effectively support social licence and how they might be improved in order to better build and maintain social licence.

DECISION-MAKING PRACTICE AND PROCESS

The book focuses on three regulatory frameworks for decision making in the context of data linkage: human rights, research ethics, and law. It then goes on to consider the practice and process of decision making within these frameworks. The focus on decision making in this book is necessary and appropriate because decision making is at the heart of administration,[5] and good decision making lies at the heart of good administration.[6] In addition, decision makers in the context of data linkage research are gatekeepers and guardians. They are the gatekeepers that regulate researcher access to the rich and valuable resources of government data collections. They are the guardians of the individual, community, and public interests involved in the endeavour. They make decisions about which data collections are linked, the way they are linked, and the way they are stored and accessed.

Research using linked data is the end result of a number of steps:

- the original collection of data;
- the establishment of linkage infrastructure; and
- the linkage, extraction, and use of data for particular research projects.

---

[4] P Carter, G Laurie, and M Dixon-Woods, 'The Social Licence for Research: Why care.data Ran into Trouble' (2015) 41(5) *Journal of Medical Ethics* 404, 404.

[5] H Simon, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations* (Macmillan, 1947).

[6] Queensland Ombudsman, 'Good Decision-Making' (Web Page, 3 June 2019) <www.ombudsman.qld.gov.au>.

Cambridge University Press & Assessment
978-1-108-42664-0 — Sharing Linked Data for Health Research
Carolyn Adams , Judy Allen , Felicity Flack
Excerpt
More Information

Each of these steps involves different stakeholders and different decision makers. The original collection of data is often for administrative – rather than research – purposes, for example the administration of public health insurance schemes and public hospitals. Such data collections generally include large numbers of people, often whole populations. The stakeholder community therefore includes the whole population. Governments must have legal authority to collect such data, and the decision makers involved in authorising that collection therefore sit in government and the legislature. In representative democracies, these decisions makers are tasked with making decisions on behalf of citizens.

In Australia, Canada, and the UK – the jurisdictions under consideration in this book – data linkage infrastructure has been established either by government or by public universities, so the stakeholder community is once again very broad. It includes the whole population as potential data donors and taxpayers, the research community, and data custodians.

The linkage and extraction of data for specific projects engage a smaller pool of data donors. Some projects may involve specific communities such as a particular disease group, a particular ethnic group, or a particular geographic or socio-demographic group. However, the impact of the project may reach beyond those whose data is included in a particular project. Other projects may include very large numbers of people spread across the whole population. The decision to share data for individual research projects is entrusted to data custodians and privacy and ethics committees.

The book considers the governance arrangements and the decisions that lead to the release of linked data for research and seeks to describe 'good decision making' in the context of linking and releasing data for research. In developing an approach to good decision making, we have drawn on the principles of public administration, which have been developed over hundreds of years to regulate government decision making about individual matters. These principles include having a sound legal basis for decisions; appointing appropriate decision makers; making decisions that are in the public interest based on explicit and publicly available criteria; and providing reasons for decisions linked to those criteria. These governance arrangements are intended to drive greater trustworthiness, transparency, accountability, and community engagement in the decision-making process. These in turn support social licence for the process and the resulting decisions.

TERMINOLOGY

There are several terms that appear throughout the book that require some clarification. The focus of the book is the use of data for research, a term that can be difficult to define. We have taken 'research' to mean the creation of new knowledge and/or the use of existing knowledge in a new and creative way so as to generate new concepts, methodologies, inventions, and understandings.

The term ‘linked data’ is used throughout the book to mean data derived from different sources but relating to the same individual or event in a single file. The method of creating linked data is called data linkage rather than other terms used in the literature, including record linkage and data integration.

The term ‘individual-level data’ is used in the book to refer to data about an individual person, whether or not the person is identified. It is contrasted with aggregated data which combines data about different people. The arguments in the book apply to all individual-level data and are not confined to personal information. The terms ‘personal information’ and ‘personal data’ are used in international human rights law and data protection legislation around the world and are defined as information about individuals that is identified or reasonably identifiable. ‘Reasonably identifiable’ is context-dependent so that information in one context may be ‘personal information’, that is, information about an individual who is reasonably identifiable, but may not be ‘personal information’ in another. Advances in the ability to link and analyse data mean that it is becoming increasingly difficult to ensure that data has been de-identified and that individuals cannot be re-identified. The terms ‘personal information’ and ‘personal data’ are only used in the book where they are descriptive of existing law, ethical principles, or procedures.

We use the term ‘data protection legislation’ to describe legislation that governs the collection, use, and disclosure of personal information by both private and government entities. Data protection legislation is also known, in different jurisdictions, as privacy legislation or personal information protection legislation.

We have chosen to use the generic term ‘ethics committee’ throughout the book for a committee that reviews research proposals involving human participants to ensure that they are ethically acceptable. We acknowledge that these committees are known by a range of terms in different countries, including Human Research Ethics Committee (HREC), Institutional Review Board (IRB), Ethics Review Board (ERB), Research Ethics Committee (REC), and Research Ethics Board (REB).

STRUCTURE OF THE BOOK

The book is divided into three parts. In Part I, the concept of data linkage is explained, the potential benefits of using linked individual-level data described, and an overview of approaches to managing and minimising privacy and other risks in different countries is included. The individual, community, and public interests in the use of linked data for research are introduced and the concept of social licence is discussed. In Part II, the three regulatory frameworks for making decisions about whether linked data can be used for research – human rights, research ethics, and law – are explored. In Part III, case studies and examples are used to investigate how these decision-making frameworks are implemented in different countries and how this might impact access to data for researchers and the social licence for research using linked data. The challenges that emerge in the implementation of decision-making processes are considered and recommendations made to enhance the regulatory and governance structures as well as support social licence.

SCOPE OF THE BOOK

While this book focuses primarily on health-related research, the concepts and arguments discussed in the book are generally applicable to other kinds of research using linked data. The challenges of regulating data use and making good decisions about the collection and use of data for research are apparent everywhere and are being addressed in different ways in different countries. The description of the legal and ethical frameworks, in Chapters 5 and 6, and the examples and case studies in Parts II and III, are drawn from common law countries that have advanced data linkage programmes: mostly Australia, Canada, and the UK. They illustrate the way in which these challenges are currently being addressed in these jurisdictions. The comparative and critical approach taken to this material will, we trust, ensure that the general concepts discussed in this book are internationally applicable and will be relevant to addressing those challenges in all jurisdictions. The law and other content in the book are up to date as of 30 June 2021.

Cambridge University Press & Assessment
978-1-108-42664-0 — Sharing Linked Data for Health Research
Carolyn Adams , Judy Allen , Felicity Flack
Excerpt
More Information

PART I

Context for Decision Making

In this book, we are interested in assessing and improving the decision-making process that grants or denies access to government data collections for research. Sound decision making is foundational to good administration, including managing public resources efficiently and effectively. The Australian, Canadian, and United Kingdom (UK) governments have all made clear that the data they hold is a resource that must be managed for the benefit of the community.[1]

Governments make a great deal of administrative data publicly available for easy access and reuse. However, this is clearly not appropriate in relation to individual-level data about intimate aspects of people's lives, such as their health information. When open access to data is not an acceptable option, then decisions must be made about each individual use and release of the data.

This book is concerned with one use of that information – linking data for research – and the way decisions are made about this use of data. These decisions are challenging because of the multiplicity of interrelated interests and the impact they have on individuals and the wider community. They engage deeply held moral values and convictions.

In Part I, the context for decision making in relation to the use of data for research is established. The current landscape of data linkage for research is described, the interests of the stakeholders are discussed, and the concepts used to analyse and evaluate the decision-making frameworks and procedures are examined.

---

[1] Australian Government, 'Public Data Policy Statement' (Policy Statement, 7 December 2015) <https://pmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf>.

In Chapter 1, the concept of data linkage is explained and the potential benefits of using linked individual-level data are described. The types of data that are the focus of the book are introduced, and an overview of the approaches to managing and minimising privacy risks adopted in different countries is included.

Chapter 2 considers the perspectives of different stakeholders in research using linked data and examines the interests of those who conduct the research, the interests of individuals whose information is used, collective interests, and public interests in the use of linked data for research. The complexity and interrelationship of these interests reveal how challenging it is to make decisions that take all of them into account. The chapter goes on to consider some concepts that shed light on the way those interests should be evaluated when decisions are made.

Chapter 3 introduces the concept of social licence and considers its application to research using linked data. Social licence is essential for the ongoing viability of research using linked data. The concerns about data linkage and the conditions that support a social licence for research using linked data are discussed. The necessary attributes of governance mechanisms and the implications for the regulation of research using linked data are examined.