BIG DATA *for* AUSTRALIAN SOCIAL POLICY

Developments, Benefits and Risks

Edited by

Janet Chan
Peter Saunders

ACADEMY OF
THE SOCIAL SCIENCES
IN AUSTRALIA

ACADEMY OF
THE SOCIAL SCIENCES
IN AUSTRALIA

Acknowledgement of Country

The Academy of the Social Sciences in Australia acknowledges the Traditional Owners and custodians on whose lands the Academy's National Office is located, the Ngunnawal and Ngambri People, and all of the Traditional Owners on whose lands the Academy undertakes its activities on which Academy Fellows live and work.

How to cite this report

Chan, J. and Saunders, P. (Eds) (2021). Big data for Australian social policy: Developments, benefits and risks. Academy of the Social Sciences in Australia, Canberra.

License

ISBN 978-0-908290-02-4
9 780908 290024 >

FOREWORD

When I led the Commonwealth Health Department 20 years ago, there was almost universal enthusiasm amongst health experts about the potential benefits of a comprehensive electronic health record (EHR). The potential benefits were considerable: a shift to a more patient-oriented health system, the ability to offer more appropriate and effective services for the growing numbers of people with chronic conditions, improvements in early diagnosis and public health interventions, greater efficiency from reduced duplication of diagnostic testing, and reduced pressures on hospitals. Some of the risks were well known, particularly around privacy, but there was naïveté about the scale of other risks, including those surrounding contested policy priorities and interests, governance arrangements, the rights of patients and doctors, and the technical challenges associated with the linked data. Twenty years on, progress has been far more limited than hoped for, as the risks and difficulties became more apparent, even as technology continued to offer more solutions and more potential gains.

In those 20 years, there have been enormous advances in the technical capacity to develop and utilise “big data” and the potential benefits involved are now extended across all areas of public policy (and across the private sector). As with the EHR, they include: a more “citizens-centred” (or “consumer-oriented”) approach; the ability to track decisions and developments over time not only to tailor and target future services but also to investigate likely causes and effects, and thereby achieve more “outcomes-focused” policies; the ability for real-time monitoring of events and developments; and the capacity to draw on comprehensive data sources rather than sample surveys and hence to analyse particular localities or groups.

As this book reveals, it is easy to become over-enthusiastic about the potential benefits of “big data.” And the risks involved need to be carefully identified and managed. Through a series of social policy case studies, including interviews with experts involved, as well as examination of associated documentation, the authors reveal whether the potential benefits were realised, where unintended adverse impacts occurred, and lessons for good practice risk management. The editors, in the book’s opening and closing chapters, draw together the issues raised and present the consensus view of how “big data” initiatives for social policy should be conducted, particularly in the Australian context.

In summary, the editors suggest that critical to successful use of big data for social policy are:

- The governance arrangements that promote trust, including transparency and security;
- Adequate infrastructure, including the skills and capacity needed to understand the quality of the data and to use it appropriately, appreciating the extent of uncertainty of any conclusions drawn;
- Human rights, recognising the sovereignty of those about whom the data is concerned (with particular reference to Indigenous people and communities);
- Consent and a “social licence” to use the data for the purposes involved; and, related to all these,
- Accountability.

This book goes into much detail about these and related issues and suggestions. It provides essential reading for government officials, academic researchers and community leaders.

- Andrew Podger

ACKNOWLEDGEMENTS

This project was funded by an Australian Research Council Linkage Learned Academies Special Projects (LASP) grant (LA170100011) with the Academy of Social Sciences in Australia as the administering organisation. The project involves the collaboration of researchers from UNSW Sydney, University of Queensland and Griffith University.

We would like to thank the project's Expert Advisory Group—Professor Lyria Bennett Moses, Professor Michele Haynes, Professor Kerrie Mengersen, Professor Julia Lane, Professor Samantha Low Choy, Professor Andrew Podger AO and Dr Dennis Trewin AO—for their valuable guidance throughout the project.

The administrative assistance from the Academy, the Social Policy Research Centre (UNSW), the Institute for Social Science Research (UQ) and Griffith University is gratefully acknowledged. We are especially grateful for the support of Dr Chris Hatherly, CEO of the Academy.

We would also like to thank the experts who took part in the interviews for Chapters 2, 4, 5 and 6, and Associate Professor Bruce Bradbury for his assistance with the establishment of the project's data archive.

We also appreciate the useful exchange of ideas over the last few years with Emeritus Professor Michael Barber AO FAA FTSE, leader of the Australian Academy of Science LASP project on Big Data, who also took part in one of our workshops.

CONTRIBUTORS

David Abramson FACM FIEEE FTSE FACS is a chief investigator of this project, Professor of Computer Science, and Director, Centre for Research Computing, University of Queensland

Janeen Baxter FASSA is a chief investigator of this project, Professor at the University of Queensland and Director of the ARC Centre of Excellence for Children and Families over the Life Course

Lyria Bennett Moses is Professor and Director, Allens Hub for Technology, Law and Innovation, Faculty of Law and Justice, UNSW Sydney

Janet Chan FASSA is co-leader of this project, Emeritus Professor and co-leader of the Data Justice Research Network, Allens Hub for Technology, Law and Innovation, Faculty of Law and Justice, UNSW Sydney

Marco Fahmi is Program Manager for Text Analytics and Language Commons at the School of Languages and Cultures at the University of Queensland

Robyn Gulliver is a Postdoctoral Research Fellow, Department of Psychology, University of Hong Kong

Ross Homel AO FASSA is a chief investigator of this project and Foundation Professor of Criminology, Griffith University

Sara Kalucza is a Research Fellow at Umeå University, previously at the Life Course Centre, University of Queensland

Ilan Katz is Professor, Social Policy Research Centre, UNSW Sydney

Samantha Low-Choy is Associate Professor and Senior Statistician, Griffith University

Sonia Pervin is a PhD candidate and Research Assistant in the Institute for Social Science Research and the Life Course Centre at the University of Queensland

Judy Rose is a Research Fellow and Mixed Methods Advisor, Griffith University

Peter Saunders FASSA is co-leader of this project and Emeritus Professor, Social Policy Research Centre, UNSW Sydney, where he was Director between 1987 and 2009

Sheila Shaver FASSA is a chief investigator of this project and Emeritus Honorary Professor, Social Policy Research Centre, UNSW Sydney

kylie valentine is Professor and Research Director, Social Policy Research Centre, UNSW Sydney

Mark Western FASSA is a chief investigator of this project, Professor and Director of the Institute for Social Science Research, University of Queensland

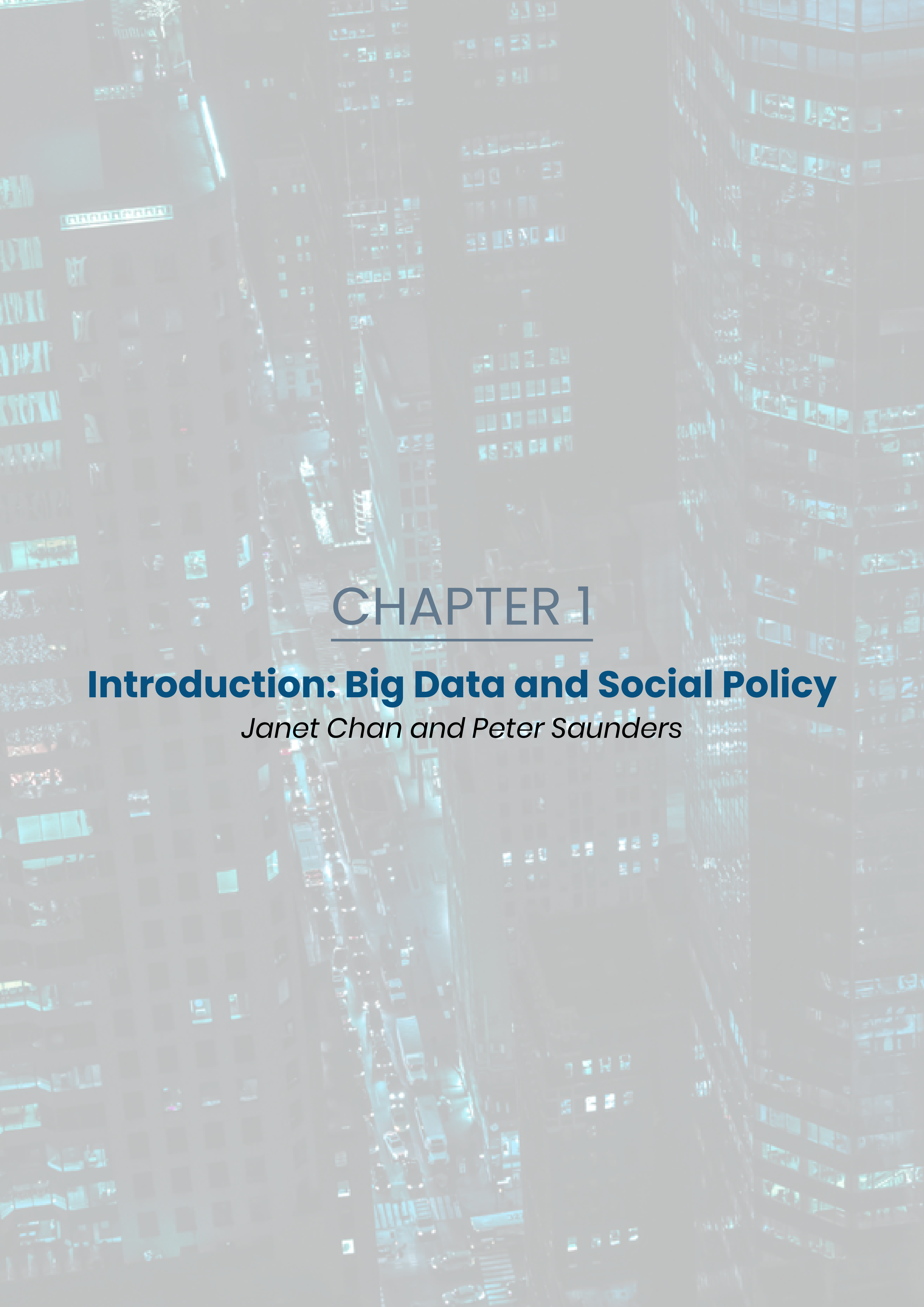
Ning Xiang is a Senior Research Assistant at the Institute for Social Science Research, University of Queensland

CONTENTS

Data Practices in a Web of Values: Reflections on the Gap between Ethical Principles and Data-Driven Social Policy
Lyria Bennett Moses, kylie valentine and Janet Chan

Technical Considerations when Implementing Digital Infrastructure for Social Policy
Robyn Gulliver, Marco Fahmi and David Abramson

Concluding Reflections
Peter Saunders and Janet Chan



CHAPTER 1

Introduction: Big Data and Social Policy

Janet Chan and Peter Saunders

Introduction: Big Data and Social Policy*

This book presents the findings of a multidisciplinary research project that investigates the benefits and risks of using big data and associated analytical tools for the design, implementation and evaluation of social policy, or for improving program delivery.

KEY CONCEPTS AND RECENT DEVELOPMENTS

The scope, coverage, quality, timeliness, sources and uses of data are expanding rapidly. So too is the ability to link data from multiple sources covering key socioeconomic variables, in aggregate, and for individuals and different socioeconomic groups. In Australia, following a report by the Productivity Commission (2017), a National Data Commissioner has been established and the Data Availability and Transparency Bill 2020 has passed into legislation (Curchin & Edwards, 2021). The Multi-Agency Data Integration Project (MADIP) was established as a secure data asset within and under the control of the Australian Bureau of Statistics (ABS) in 2015 and its development is being guided by the Data Integration Partnership for Australia (see https://www.abs.gov.au/about/data-sevices/data-integration/integrated-data/multi-agency-data-integration-project-madip).

These important developments are being driven by the agencies of government that have a direct interest in the potential benefits from expanding and linking available administrative and other forms of data. These benefits are expected to accrue to all spheres of government activity, including in social policy. How specific changes are formulated, debated, assessed, implemented, monitored and evaluated has the potential to improve but also impede how programs are administered and function. Social policies and programs exert multiple effects on the lives of what are often vulnerable individuals, families and groups. Negative impacts and the factors that facilitate them also need to be considered, and, wherever possible, avoided. There are benefits but there are also risks. Both need to be identified, analysed and balanced in order to ensure that big data initiatives have positive effects, not only for the government and other agencies that employ them but also for those that depend on the services and assistance provided.

"Big data" is an umbrella term that covers the many dimensions of the underlying idea, which De Mauro et al. (2016) group into four overlapping themes: information, technology, methods and impact. After reviewing the many specific definitions, they propose the following "consensual" definition (2016: 131):

> *Big data is the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.*

This definition emphasises the three key elements of big data—volume, velocity and variety—that present the challenges and opportunities that big data unleashes, whilst emphasising that its ultimate purpose is to add value.

Southerton (2022, forthcoming) defines the related process of datafication as referring:

> *… to a process by which subjects, objects and practices are transformed into digital data … Datafication renders a diverse range of information as machine-readable, quantifiable data for the purpose of aggregation and analysis [and] is also used as a term to describe a logic that sees things in the world as sources of data to be "mined" for correlations or sold, and from which insights can be gained about human behaviour and social issues.*

Big data and datafication are closely related terms, and both have their uses and are employed throughout this book to refer to the same underlying phenomenon. One example is the application of artificial intelligence (AI), defined in a recent report by the Australian Council of Learned Academies (ACOLA) as: "a collection of interrelated technologies used to solve problems that would otherwise require human cognition" (Walsh et al., 2019: 2)—but there are many other definitions.

The application of big data to social policy has the potential to fundamentally alter prevailing understandings about the nature of social problems. The many benefits include drawing on a range of data sources to map social problems "on the ground," giving greater emphasis to those affected by social policies (including small-sized, hard-to-reach groups) through more sophisticated modelling and closer real-time monitoring of the impacts (intended and unintended) of policy interventions, and

---

** The authors acknowledge comments provided by Andrew Podger and Dennis Trewin on the original draft, but retain responsibility for the views expressed.*

better coordination of disparate policies. These benefits offer the promise of enhancing the wellbeing of vulnerable populations, but also embody associated risks, including those associated with poor data quality, inappropriate use of automation and related techniques, lack of effective governance and transparency, and inadequate attention paid to the specific issues faced by vulnerable individuals and groups. These benefits and risks must be balanced when deciding the specific where, how, when and to whom big data initiatives are applied, while taking account of the relevant contexts—political and social as well as economic and technical.

As the ACOLA report notes at the outset:

> *… to ensure that AI technologies provide equitable opportunities, foster social inclusion and distribute advantages throughout every sector of society, it will be necessary to develop AI in accordance with broader social, societal principles centred on improving prosperity, addressing inequity and continued betterment. (Walsh et al., 2019: 2)*

The impact of big data developments in the social sciences has been emphasised by the editors of a recent Special Issue of the leading international journal Social Indicators Research, on Data Science and its Applications to Social Research. They note in their introduction that:

> *We are facing a data revolution that is transforming the routines of production of data, once consolidated within the different disciplines. These changes are having important consequences such as the re-emergence of “data-driven” science, which is opposed to “hypothesis-driven” science. The availability of big data is facilitating the adoption of the “data first” model of science affirming that exploring existing data may be more useful than building models of why people behave the way they do. (Crocetta et al., 2021: 339)*

Big data is thus not simply affecting how social policy is designed and implemented but also the ways in which the evidence used to support such interventions is gathered, manipulated, analysed and interpreted.

A number of social science studies have cast doubt on whether progress to date has been positive overall, or whether the opportunities associated with the application of big data in the social policy sphere can ever be realised without fundamental changes in the rationale and practice of its implementation. In their introduction to a Special Issue of The Australian Journal of Social Issues devoted to big data and social policy in Australia, journal co-editors Katherine Curchin and Ben Edwards draw attention to the central role of government agencies in Australian developments. They note the secrecy surrounding many of their inputs (including, for example, the description of government agency projects undertaken under the MADIP initiative) but go on to emphasise how this lack of transparency can undermine public trust and the social licence on which the legitimacy of big data initiatives ultimately depends. They conclude their overview by noting:

> *We support greater transparency about what the government is doing with its data assets. The Australian social research community has an important and urgent role to play in uncovering the social costs and potential public benefits of collecting, storing, linking and analysing big data … The rights and interests of Australia’s most disempowered people should be central to the discussion about the data which is generated by encounters between the state and people in moments of suffering and vulnerability. (Curchin & Edwards, 2021: 139-40)*

These sentiments have been voiced by other social scientists when analysing the nature and impacts of big data and datafication. For example, Dencik and Kaun (2020: 2) emphasise the need to:

> *… explore the multiple ways in which the practices, values and logics that underpin the advancement of datafication intersect with the practices, values and logics that form the basis of the public services that we commonly associate with the modern welfare state.*

This is a major endeavour that is still in its infancy, although the rapidly expanding number of practical examples of datafication provide fertile ground to begin to assess some of the challenges raised and explore how well their impacts align with what was promised. The chapters in this book are designed to illustrate the variety of applications already present in social policy, and identify some of the key themes and issues that will shape how future big data developments can better acknowledge, analyse and address the associated benefits and risks.

THE AUSTRALIAN CONTEXT

At first glance, Australia has much to gain from the application of big data to its social policy settings. The defining feature of the Australian approach (particularly in relation to its income support programs) is its focus on targeting assistance to those in greatest need. This provides an imperative to draw on big data developments because information about individual circumstances is a key determinant of who receives assistance, under what conditions, in what form and how much. The absence of any form of earmarking of contributions—a central feature of the social insurance schemes that are common in Europe and (in selected form) in Asia—places greater reliance in Australia on the ability of general tax revenue to fund social programs, and this, in turn, raises questions about their financial and political sustainability. Big data can play a role in generating insights into the dynamics of movements in and out of targeted categories and help identify interventions that can promote positive movements and improve people's independence and wellbeing. But there are many risks, particularly when the primary goal is to restrain expenditure growth or cut costs for fiscal reasons, where the burden often falls mainly on those most vulnerable and least able to adjust.

In this context, the use of performance management to promote "results based" administration (Podger, 2020) provides ample scope for new forms of data to underpin improvements in efficiency and performance, and inform more robust evaluations, but only if the associated risks are acknowledged and managed effectively. These risks generally fall on those at the receiving end of policy, and, if not addressed, can compound the problems associated with the interconnectedness and complexity of many social problems. As these forces place upward pressure on social budgets, the political response has been to "do more with less" in order to avoid raising taxes—a challenge that remains formidable despite several decades of spending restraint.

There is thus a gap—arguably a gulf—between the laudable policy objectives identified by bodies like ACOLA and the political and ideological priorities driving policy, particularly social policy. This gap has many dimensions, reflecting the multiple ways in which the datafication of society introduces and entrenches key questions relating to issues of inequality, discrimination and exclusion that impinge on the broader debate about social justice (Dencik et al., 2018: 1). Far from being just a technical advance, the shift to automation in determining what counts as social knowledge raises issues about the meaning and achievement of social justice itself, and hence about the shape and content of the policies put in place to achieve it. Dencik et al. argue that the starting point should focus on claims of injustice, since this provides a gateway to examining the source of these claims rather than being distracted by what justice itself looks like. This approach provides an avenue for exploring how datafication creates new barriers to inclusion if those on whom it is practised are not fully consulted and involved in all stages of its development and application. Some have argued that datafication is part of a political agenda being used by governments to promise technological solutions to social problems, but this agenda must be resisted by reframing big data as a social justice concern.

These issues have been explored in detail by Dencik and Kaun (2020) against the backdrop of a welfare state that is being transformed by governments employing privatisation and other public management initiatives in the search for greater efficiency. Big data and datafication are playing a key role in this process—not surprisingly, since the creation of databases and the monitoring of citizens have been part of the architecture required to determine population needs and guide resource allocation decisions. However, recent developments have increased the tension between an informed and overbearing state and the rights of citizens, as discussed in a recent report by the UN special rapporteur on extreme poverty and human rights (Alston, 2019). The report argues that systems of social protection and assistance are being increasingly driven by the use of digital data technologies to "automate, predict, identify, surveil, detect, target and punish" (Alston, 2019, cited in Dencik & Kaun, 2020: 2). Citizens are being given little opportunity to participate in the choices and values embedded in these developments, the process instead being led by government as a political project, not just a technological one.

The resulting lack of scrutiny raises concerns about transparency and fair treatment, while further isolating citizens from a process that is increasingly impinging on their lives. For example, the use of "citizen scoring" to predict the likelihood of future acts, including committing a crime, becoming a teenage mother or experiencing long-term unemployment raises a raft of concerns when it is used to justify early interventions "designed" to prevent these future events. The mathematical calculus may well suggest future savings in public expenditure, but only when aggregate probabilities are assigned to individuals and then used to deny benefits or impose sanctions now. One danger of this approach is its focus on individual characteristics as the target of remedial action, to the neglect of structural factors. These features raise important questions for social policy that need to be exposed, analysed, researched and debated, but the cloak of secrecy surrounding many aspects of big data impedes efforts to connect policy "responses" to individual experiences.

Some of the consequences of these developments are explored in the Australian context by James and Whelan (2021). They draw the contrast between how the concept of "ethical AI" is portrayed in government documents and other high-level reports and how it is perceived by those at the receiving end, for whom:

> *Digitisation frequently appears to entail inappropriate encroachment on personal data, botched in execution, to ends contrary to public best interest (James and Whelan, 2021: 10).*

They go on to discuss a range of alleged "botch-ups" before discussing the Robodebt controversy in depth, drawing on analyses conducted by others, including Cao (2012), Carney (2018, 2019), Henman, (2017, 2019), Macleod (2017) and Mann (2019).

Robodebt laid bare the problems that can arise when computer algorithms are allowed to directly affect program delivery without adequate human overview. James and Whelan reach two general conclusions: first, that ethics discussions do not challenge the legitimacy of AI but are urgently needed, given that AI is imminent and inescapable; and secondly, that the dynamics of AI development are driven by those within powerful institutional systems who are able to influence the identification of the "problems" to which it is applied.

Robodebt is discussed in detail later (in Chapter 4) but it is relevant to acknowledge here that data matching represents a continuation of past practices, not a radical departure. Data matching has long been a cornerstone of efforts by the (former) Department of Social Security (DSS) to identify potential cases of fraud or over-payment. The Data-Matching Program (Assistance and Tax) Act 1990 formalised data matching as a key component of internal administration of the social security system to ensure compliance with legislated entitlements. Following the introduction of the Act, departmental procedures were amended, with the DSS 1992-1993 Annual Report noting that:

> *The department has revised all its procedures for data-matching exercises involving external data. This ensures that in all but exceptional cases clients who are subject of an apparent match are contacted first to verify information before a third party, such as an employer, is contacted. (DSS, 1993: 80)*

What changed with Robodebt was not the imperative to identify and combat fraud and incorrect payment but its application. Previously, data matching was used to identify potential "discrepancies," that were then reviewed by a departmental officer and only then triggered the sending of a letter seeking an explanation. Advances in automation made it possible for letters to be sent out automatically to all discrepant cases. In the process, the element of human oversight was dispensed with entirely, and those identified as potential problem cases previously were now assumed to be "guilty as charged" and required to repay the "debt" incurred unless able to produce evidence to the contrary. The balance shifted from a presumption of innocence to a presumption of guilt, with the complex and cumbersome online appeals system serving to discourage those who wished to dispute their guilt.

The shift represented by Robodebt was important and was facilitated by technological developments, but the driving force was an unrealistic appreciation of what automation could achieve as a conduit linking the state to the recipients of its programs. As Whiteford (2021: 356) points out, the failure of the Robodebt "fiasco" was primarily caused by the unlawful application of income averaging and a wilful disregard of the adverse effects imposed on "hundreds of thousands of members of the public." A more cautious approach that drew on past experience and contained appropriate review and safety mechanisms could have achieved much of what the government claimed was its goal, without imposing fear and hardship on thousands of vulnerable clients, not to mention the humiliating backdown that Robodebt eventually produced.

SOCIAL POLICY AND THE WELFARE STATE

Proceeding in parallel with these consequences of the datafication process is a literature on the role of big data in managing the evolution of many welfare state programs. These pressures reflect concern (often driven more by ideological obsession than hard evidence) that the disincentive effects of state intervention will induce behavioural changes that undermine economic performance. Much depends on how policies are framed and on the degree to which different groups respond to price signals and how this is reflected in the way policy is designed and enacted. These factors help explain why the benefits of "smart" or "big" data and the associated analytics have been extensively promoted in a number of areas, including predicting crime and recidivism (Bennett Moses & Chan, 2014, 2018), assessing the "value added" by schools and teachers (Amrein-Beardsely, 2014), and distinguishing window shoppers from buyers on commercial websites (O'Neill, 2016).

Policy became increasingly based on the premise that since the use of algorithms to identify instances of fraud could be made, it should be made—irrespective of any broader calculation of its overall costs and benefits. This narrow focus on the rationale for the implementation of automation and algorithms in program design and delivery explains how and why the process was allowed to cause serious harm to many, who were required to establish their innocence of what the automated system had identified as a breach.

One goal of this book is to redress these unwarranted (in terms of their rationale) and unjust (in terms of their impact) interventions and broaden the debate to include the application of big data and datafication in other key areas of social policy, and to draw on international as well as Australian experience.

The focus to this point on social security issues reflects a combination of cost and historical experience. In terms of cost, the social security budget in 2019-20 was $196.1 million, equivalent to just under 10% of GDP and accounting for one-third (33.9%) of total Commonwealth budget outlays of $578.5 million, far outstripping spending in any other area (Australian Government, 2020). Substantial amounts are spent by the Commonwealth on other areas of the welfare state, primarily on health, education and aged, disability and child care services, and these amounts are supplemented by substantial spending by state and territory governments. These programs mainly provide services rather than income support, but reducing spending has proved difficult because the scope for productivity improvements is limited by the labour intensive nature of most human services.

While there is scope to better target resources in service delivery, it is mainly in relation to social security that effort has been concentrated, eliminating "waste" and improving efficiency in order to relieve the pressure on spending and its growth. The view that government programs were creating a culture of "welfare dependency" emerged under the Howard government. Given the political difficulty of cutting benefits, the shift of focus onto combating such dependency provided the backdrop to a new narrative that placed the onus on reducing recipient numbers, particularly where this could be achieved by restricting or denying benefits to those who were exploiting the system. There was little evidence to support such claims and the predicted "savings" were wildly exaggerated and thus largely illusory, but thus began the logic that led to Robodebt.

THE ROLE OF BIG DATA

The application of big data techniques, when conducted appropriately, is consistent with broader government aims to improve program efficiency and minimise costs by better monitoring (and/or increased surveillance) of those applying for and receiving assistance. These ideas are not new, but their attainment has taken on a new dynamic as the scope of datafication, including through data linkage and data matching, has expanded in scope and intensity. Recent social security big data initiatives are supporting aspects of the Australian Government's Priority Investment Approach to Welfare (PWC, 2016), which develops a synthetic population model to assess lifetime liabilities and targets cohorts of welfare recipients for program attention. But as Bray et al. (2020) have observed, these initiatives have also highlighted the growing tension between searching objectively for better, evidence-based policies and the often subjective and highly political nature of government decisions.

Australian experience with the use of big data in the design and implementation of social programs has highlighted a need for a cautious approach that recognises the importance of setting realistic expectations about what can be achieved. These concerns have been expressed by international scholars in relation to the use of big data analytics for general or specific applications (see boyd & Crawford, 2012; O'Neill, 2016). Big data offers considerable promise for better and richer descriptions of social problems, for better targeted and coordinated policy responses and for improvements in policy and program evaluations, including in areas where statistical methods do not perform well, such as "out-of-sample" prediction (Kleinberg et al., 2015). However, the approaches can also have adverse effects on vulnerable populations by masking crucial assumptions and undermining democratic decision-making (O'Neill, 2016). Mistakes can have dire consequences for some of the most vulnerable in society, yet these negative effects rarely enter into the calculus that supports the use of big data or are often greatly underestimated when they are.

These issues have been given prominence by the OECD (2019) which has identified key issues and reviewed selective member country initiatives in big data in order to promote the sharing of experiences and responses. It identifies the key role of social protection systems in mitigating the risks facing individuals and cushioning the interactions between them that can deepen the flow-on consequences (for example, of job loss leading to poor health, of relationship break-up leading to poverty, or of lack of engagement with the schooling system leading to youth unemployment). These consequences can result in escalating social policy costs, but when applied appropriately, the "smart use of data can contribute to a more effective social policy that makes it less likely that such a situation arises" (OECD, 2019: 1).

However, data does not exist in a vacuum: it must be specified, collected, assembled, stored, validated, documented, shared, analysed and interpreted. Each step in this process raises important ethical issues relating to consent, legality, transparency, fairness, privacy and access. There is also the overriding issue of data quality, which takes on a new dimension when data from different sources, collected for different purposes by different organisations, and with different relations to real time, are linked or matched. Interpreting analytical results requires some understanding of the accuracy of the individual data and matching variables and algorithms, but also requires expertise and judgement that may mitigate drawing firm policy conclusions, particularly given the sensitivities involved.

The OECD report identifies several factors that need to be addressed before the benefits of big data can be realised and the risks avoided, or at least minimised. It concludes that the application of data matching and related interventions should be subject to the same processes that underpin sound social policy practice in other areas, covering the four phases of the policy cycle identified by Studinka and Guenduez (2018): planning, design, delivery and evaluation.

To date, few Australian researchers have actively investigated the application of big data tools to social policy (a notable exception is the work of Henman, 2010, 2019, 2020). There is an urgent need to promote increased understanding of the technical aspects of the analytical tools associated with datafication, draw out the assumptions underlying the modelling, highlight the often wide gap between the assumptions that underpin the modelling and the resulting policy interventions, and take better account of the (intended and unintended) consequences of these applications, particularly for vulnerable groups in the community. This requires cross-disciplinary collaboration in researching how data analytics can be used to guide social policy in a constructive and appropriate way, including by identifying the risks and pitfalls to guard against.

OUTLINE OF THE BOOK

In assembling some of the evidence needed to achieve the goals identified above, the chapters in this book examine the role of big data in a variety of social policy settings and by drawing on specific examples of national and international experience. Some chapters present case studies of specific issues, policies or experiences while others report broader findings using a variety of qualitative and quantitative techniques, or draw on recent evidence to identify key themes, implications or (tentative) conclusions. The issues addressed examine Australian and international experience using a variety of methodologies and analytical techniques on different forms of data to reveal important insights and highlight the many chains of influence linking data developments, policy goals and social impacts.

Chapters 2 and 3 describe international experience, while Chapters 4 to 6 turn to Australian examples of the application of big data in a variety of social policy contexts. Although differing in their content and approach, the chapters all focus on identifying the benefits and risks of using big data and datafication analytics such as data matching, machine learning and similar tools for decision making in the design, implementation and evaluation of a range of social policies. They present a series of examples that draw on the available literature and policy documents, and in-depth interviews conducted with key participants in selected big data initiatives. The aim is to better understand the motivations of those involved, map outcomes against anticipated achievements and identify the key lessons for future applications. Chapters 7 and 8 address broad themes in the application of big data to social policy, the first relating to ethical issues that emerge from the earlier case study chapters, and the second addressing the key issue of infrastructure. The final chapter draws together some reflective conclusions.

The forces driving the developments discussed throughout the book are too broad for their many implications, consequences, benefits and risks to be covered comprehensively in a single volume. Despite this, the perspectives, approaches, techniques, findings and implications contribute new perspectives on Australian social policy and highlight areas and issues that will require ongoing attention.

The broad content of the book and the individual chapters were developed, discussed and refined in three workshops attended by the project investigators and members of an Expert Advisory Group (EAG) appointed to review and guide key elements of the project. Data collection instruments were designed and discussed at an initial workshop, and ethics approval was obtained from the three participating universities. Preliminary results were presented to two further workshops attended by EAG members and selected external experts, who also provided comment and advice. Additional feedback was provided by participants in the 2019 Australian Social Policy Conference, at which several papers were presented.

SUMMARY OF CHAPTERS

Chapter 2, “Social Investment: The New Zealand Case,” describes the history of the New Zealand social investment approach, focusing on the opportunities and challenges in relation to the use of government administrative data for actuarial analysis and program development. The approach was introduced in 2010/11 by the former National-led Government, with the aim of addressing long-term welfare dependency and the increasing fiscal liabilities of the welfare system. The core elements of the approach were threefold: to harness the analytical and actuarial power of large-scale linked government administrative data to identify groups at risk of long-term social welfare dependence, to tailor social interventions to the needs of specific social groups and clients and design programs that showed clear returns on investment designed to build human capital, and to work collaboratively across government and non-government agencies and sectors to develop and implement social services. The approach has undergone several modifications as it has evolved. Drawing on a review of key documents and interviews with five participants, the chapter highlights the importance of a shared vision, clarity of core concepts, adequate governance and infrastructure, as well as political support for such an approach to realise its objectives.

The focus of Chapter 3, “Value-Added Models in Education Performance Management,” is on the use of ongoing standardised achievement testing and value-added statistical models by US policymakers to monitor school and teacher performance and determine the value they add to student achievement. These methods have been accompanied by strong accountability regimes which reward and punish states, schools, school districts, teachers and students for perceived good or bad performance. The chapter discusses the theory behind Value-Added Models, their implementation in different jurisdictions and the legal, ethical and policy challenges of using such models for high-stakes decisions. The case study has important implications for the use of big data for social policy, such as the need to understand the operational context, consider implementation challenges and establish procedures that are transparent, accountable and contestable.

The next three chapters address a series of interrelated themes in the context of three important social policy initiatives by drawing on experience with a variety of data linkage and data-matching initiatives. The evidence presented draws on material collected through interviews with experts in a range of senior positions involved in the development of social policy and programs, within government and non-government agencies. The core interviews are supplemented with two that addressed social security issues specifically and focused on how big data influenced the evolution of the administration of the social security system prior to and during the Robodebt experience. Further details of those interviewed and detailed descriptions and analysis of the views expressed are described in each chapter.

Chapter 4, “Big Data and Government Services: Threats and Opportunities for Data Linkage and Subsequent Analytics” is the first of three chapters that explores, through interviews, how experts viewed the current state of big data/data linkage in Australia. This chapter focuses on the role of technical capability, skills and infrastructure. From this angle, social and ethical issues emerged regarding intended use of data linkage, or on the misuse of data linkage. The discussion examines what lessons were learned and are being assimilated, from the Robodebt controversy, given that these experts were interviewed before the illegality of Robodebt was formally recognized. Three common pitfalls were identified when data linkage occurs within a governance context. Data linkage may be problematic when used for compliance. Accountability is best embedded within technical efforts. Social licence precedes data availability. Four lessons highlight misunderstandings. Problems in one facet of data linkage are often misconstrued as limitations in another. Successes and failures of data analytics may be mis/attributed to data linkage. Misuse is not necessarily technical, instead relating to communication, transparency, accountability and legality. Supply and analysis of linked data may better address social/ethical issues if treated as official statistics. Finally, this suggests a more multi-disciplinary and cross-sectoral approach.

Chapter 5, “Strategic Data Linkage to Improve the Wellbeing of Vulnerable Children: Reflections of Experts,” examines how data linkage and integration promise significant improvements in the relevance, immediacy, comprehensiveness, and hence the social value of big datasets. It argues that there needs to be a balance between technical feasibility, community trust, policy and practice sensibilities, ethical imperatives in terms of privacy, and issues of data sovereignty. This is especially the case when data relate to vulnerable children. The chapter reports on interviews conducted with 10 influential experts from health, education, human rights and child and family welfare statistics sectors. Several interrelated themes emerged from the data, relating to: data-sharing legislation and infrastructure (skills, capacity, resources); governance (ethical issues and privacy, security, trust); human rights, including Indigenous data sovereignty; and consent from data subjects and the establishment of a social licence. Some experts viewed these issues as hurdles to be overcome by organisations and government, while others proposed community consultative approaches to address complex problems such as linking administrative and child wellbeing data to evaluate place-based interventions. The major insight afforded by the analyses is that the technical, governance, and rights domains are intimately connected, with reforms in one domain having significant implications for other domains. This holistic perspective is one of the advantages of expert views.

Chapter 6, “Data Linkage and the Evaluation of the Impact of Place-Based Initiatives for Families and Children,” focuses on data and methodological issues associated with the evaluation of place-based initiatives (PBIs) in social policy. The chapter describes the advantages of PBIs over traditional preventive or early intervention approaches to improve the wellbeing of vulnerable children, and goes on to outline the methodological and practical challenges for conducting robust evaluations of these initiatives. While advances in data linkage technology and analytics provide a potential solution to resolving these challenges, significant technical, ethical and governance issues remain, as highlighted by experts who were interviewed. The authors conclude that progress will require national leadership, the development of technical expertise and policy/practice understanding and early engagement with community, data custodians and stakeholders to build trust in the design and implementation of such data linkage projects.

Chapter 7, “Data Practices in a Web of Values: Reflections on the Gap between Ethical Principles and Data-Driven Social Policy,” examines the ethical dimensions of the datafication of social policy. Using the OECD’s Principles on Artificial Intelligence as an analytic framework, and referring to the findings of Chapters 2 to 6, the chapter discusses the extent to which ethical principles (such as inclusive growth, sustainable development and wellbeing; human-centred values and fairness, transparency and explainability, robustness and accountability) have informed the design of these programs/approaches or are reflected in their implementation. The authors explain the gap between principles and practice in terms of the context in which social policy programs are formulated and implemented: the technoscientific optimism associated with datafication must confront the pre-existing and often longstanding concerns in the formulation and implementation of social policy, e.g., adequacy of resource allocation (including technical resources), issues of social licence (including trust and privacy), and impact on disadvantaged or marginalised groups. Ethical principles formulated in response to specific types of technological advances such as datafication, though necessary, are unlikely to overcome the cultural and political barriers to the delivery of fair and accountable outcomes in social policy.

Chapter 8, “Technical Considerations when Implementing Digital Infrastructure for Social Policy,” presents a model for evaluating and designing infrastructure for the use of big data in social policy. The model identifies key technical considerations for six stages of a data analysis pipeline: data storage, data integration, data access, data analysis, data interpretation and data operationalisation. It highlights again the important role (identified in Chapter 5) that initiatives like the “Five Safes” approach can have in setting in place a framework that can guide action, support participation and build trust. The overall model aims to assist policymakers in balancing the risks and benefits associated with different technical infrastructural choices. The usefulness of the model is demonstrated using two examples: the US E-Verify work permit system and the Australian COVIDSafe app.

Finally, Chapter 9 highlights some of the key findings that emerge from the chapters and draws together important themes that can form the basis of future work in the field.

REFERENCES

Alston, P. (2019). Climate change and poverty: Report of the Special Rapporteur on Extreme Poverty and Human Rights: Geneva: United Nations.

Amrein-Beardsley, A. (2014). Rethinking value-added models in education. New York: Routledge.

Australian Government (2020). Budget strategy and outlook. Budget Paper No. 1: 2020-21, Canberra: Commonwealth of Australia.

Bennett Moses, L., & Chan, J. (2014). Using big data for legal and law enforcement decisions. UNSW Law Journal, 37(2), 643-678.

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. Policing and Society, 28(7), 806-822.

boyd, d., & Crawford, K. (2012). Critical questions for big data. Information, Communication and Society, 15(5), 662-679.

Bray, R., Gray, M., & Stanton, D. (2020). Performance management and evaluation meets culture and politics: Australia’s experience. Asia Pacific Journal of Public Administration, 42(4), 290-302.

Cao, L (2012). Social security and social welfare data mining: An overview. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 42(6), 837-853.

Carney, T. (2018). The new digital future for welfare: Debts without legal proofs or moral authority? UNSW Law Journal Forum, 1-16. http://www.unswlawjournal.unsw.edu.au/the-forum

Carney, T. (2019). Robo-debt illegality: The seven veils of failed guarantees of the rule of law? Alternative Law Journal, 44(1), 4-10.

Crocetta, C., Carpita, M., & Perchinunno, P. (2021). Data science and its application to social research. Social Indicators Research, 156, 339-340.

Curchin, K., & Edwards, B. (2021). Introduction to the special issue on big data and social policy in Australia. Australian Journal of Social Issues, 56(2), 132-142.

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. Library Review, 65(3), 122-135.

Dencik, L., Hintz, A., Redden, J., & Warne, H. (2018). Data scores as governance: Investigating uses of citizen scoring in public services. Project Report. Cardiff University. https://d atajustice.files.wordpress.com/2018/12/data-scores-a s-governance-project-report2.pdf

Dencik, L., & Kaun, A. (2020). Datafication and the welfare state: An introduction. Global Perspectives, 1(1), 12912 https://doi.org/10.1525/gp.2020.12912

Department of Social Security (1993), Annual report, 1992-93. Canberra: DSS.

Henman, P. (2010). Governing electronically: E-government and the reconfiguration of public administration, policy, and power. Basingstoke, U.K. & New York, U.S.: Palgrave Macmillan. https://doi.org/10.1057/9780230248496

Henman, P. (2017). The computer says 'debt': Towards a critical sociology of algorithms and algorithmic governance. Zenodo. http://doi.org/10.5281/zenodo.884117

Henman, P. (2019). Automating inequality: How high-tech tools profile, police, and punish the poor. Critical Social Policy, 39(1), 151-153. https://doi.org/10.1177/0261018318808405b

Henman, P. (2020). Improving public services using artificial intelligence: Possibilities, pitfalls, governance. Asia Pacific Journal of Public Administration, 42(4), 209-221.

James, A., & Whelan, A. (2021). Ethical artificial intelligence in the welfare state: Discourse and discrepancy in Australian social services. Critical Social Policy. https://doi-org.wwwproxy1.library.unsw.edu.au/10.1177/0261018320985463

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. American Economic Review, 105, 491-495.

Macleod, L (2017). Lessons learnt about digital transformation and public administration: Centrelink's online compliance intervention. AIAL Forum, 89, 59-68.

Mann, M. (2019). Social (in)security and social (in)justice: Automation in the Australian welfare system. Global Information Society Watch 2019. Artificial Intelligence: Human Rights, Social Justice and Development series, Association for Progressive Communications (APC): pp. 68-72. https://dro.deakin.edu.au/eserv/DU:30143270/mann-socialinsecurity-2019.pdf

OECD (2019). Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies. Paris: OECD Publishing. https://doi.org/10.1787/276aaca8-en

O'Neill, C. (2016). Weapons of math destruction: How big data increases inequality, and threatens democracy. London: Allen Lane.

Podger A. (2020). Public sector use of new technology: Opportunities and challenges. Asia Pacific Journal of Public Administration, 42(4), 207-08.

Productivity Commission (2017). Data availability and use. Report No. 82, Canberra: Productivity Commission.

PWC (2016). Valuation report: Baseline valuation 30 June 2015. Report prepared for the Department of Social Services. Price-Waterhouse-Coopers, https://www.dss.gov.au/sites/default/files/documents/09_2016/baseline_valuation_results_report_accessible_version_12_july_2016_2pwc._2.pdf

Southerton, C. (in press). Datafication. In L. A. Schintler & C. L. McNeely (Eds.), Encyclopedia of big data. New York: Springer.

Studinka, J., & Guenduez, A. A. (2018). The use of big data in the public policy process: Paving the way for evidence-based governance. Presented to the European Group for Public Administration (EGPA) Conference, Lausanne. https://www.alexandria.unisg.ch/255680/

Walsh, T., Levy, N., Bell, G., Elliott, A., Maclaurin, J., Mareels, I.M.Y., & Wood, F.M. (2019). The effective and ethical development of artificial intelligence: An opportunity to improve our wellbeing. Report for the Australian Council of Learned Academies. Canberra: ACOLA. https://acola.org/wp-content/uploads/2019/07/hs4_artificial-intelligence-report.pdf

Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco—or was it something worse? Australian Journal of Public Administration, 80(2), 340-360. https://doi.org/10.1111/1467-8500.12479

CHAPTER 2

Social Investment: The New Zealand Case

Sara Kalucza and Janeen Baxter

Social Investment: The New Zealand Case

INTRODUCTION

A social investment approach to welfare may be defined in a number of ways, but at its most basic, it means the allocation of resources to programs or interventions to promote individual capabilities, which, in turn, are expected to benefit families and communities more broadly (Midgley et al., 2017). While social investment approaches to welfare based on big data analytics are a relatively recent development, the ideas underlying social investment have a long history, albeit with considerable variation in how they are conceived and implemented (Nolan, 2013; Midgley et al., 2017; Staines et al., 2020). The underlying assumption of a social investment approach to welfare is that early interventions to improve outcomes at one stage of the life course, such as through educational attainment or employment, will result in broader flow-on improvements later in the life course. Furthermore, improvements for any one individual may benefit those around them, such as children, families and communities, including potentially, across generations, through improved family resources and dynamics (Gluckman, 2017). The elements of a social investment approach focusing on the development of human capabilities (Sen, 2003), the importance of early interventions (Heckman & Cunha, 2010) and life course trajectories (Mayer, 2004; Elder, 1994) have a long history in the social sciences. What is new is the increasing availability and use of large-scale, multidimensional, longitudinal, linked datasets, often created by governments through the delivery of services and payments, along with recent advances in data science methods and techniques, that enable policymakers to potentially track and assess the outcomes of social interventions.

Some argue that social investment approaches based on big data should be treated with extreme caution, given the possibilities for welfare surveillance, threats to privacy and labelling and stigmatising vulnerable groups (Staines et al., 2020). Others offer a more moderate view, acknowledging the opportunities, limitations and challenges of social investment approaches based on big data analytics (Gluckman, 2017; Boston & Gill, 2017). Gluckman, for example, identifies challenges around the social licence and transparency of data, the need for good governance and data management, clarity of system architecture and the importance of high-quality data. These themes resonate closely with comments made by our interview participants, as will be seen below.

We draw on this background literature and information collected from analyses of policy documents and interviews with key informants to provide an overview of the New Zealand social investment approach to welfare, introduced in 2011-12 by the National Party. A key element of the social investment approach in New Zealand is actuarial analyses of integrated government administrative data to identify social groups at risk of long-term welfare dependence and to forecast their future liability. The Integrated Data Infrastructure (IDI) developed by Statistics New Zealand and comprising whole-of-population administrative data from several government departments linked to the census and several other surveys is a core component of the New Zealand approach, and will be discussed in several sections of the chapter.

We commence by briefly outlining the development and implementation of the social investment approach in the New Zealand context. We then present our case study method and describe our data collection and analytical procedures. We present our findings in relation to several broad themes emerging from our analyses. These include an outline of the variations across sectors in understanding what social investment means, an overview of key concerns raised by stakeholders about the social investment implementation process, and issues about data competency, infrastructure and social licence. We conclude with a discussion of lessons learned and implications for Australian social policy which overall point to the critical importance of governance arrangements, engagement with communities, and some of the limits of administrative data for enabling social investment to address social disadvantage.

BACKGROUND: HISTORY AND DEVELOPMENT

The New Zealand social investment approach developed during a time of wider social welfare reforms to the New Zealand welfare benefit system, including rethinking spending priorities and increased focus on improving protection systems for children (Destremau & Wilson, 2017). Table 2.1 provides a timeline of some of the key milestone dates. Factors identified as leading to the development of the approach in New Zealand included efforts to develop tighter fiscal policies to manage spending and reduce government costs, increased emphasis on more coordinated and client-focused social services and cross-agency collaboration, concerns about increased levels of long-term and intergenerational welfare dependency, and concerns about high rates of child abuse and neglect (Boston & Gill, 2017). Importantly, given the theme of the current volume, increased recognition of the value, availability and use of administrative data and the use of data analytics to predict welfare risk and evaluate the outcomes of policy changes also contributed to the new policy approach (Boston & Gill, 2017).

Boston & Gill (2017) divide the development of the New Zealand social investment approach into two stages or models. Model 1 emerged during 2011-12, with a primary focus on improving the welfare system and reducing the long-term fiscal liability of working-age beneficiaries. Under this model, reduced fiscal liability of the benefit system was considered a proxy for overall improvement in societal outcomes, with the assumption that an individual exiting from the welfare benefits system was entering the labour market (Boston & Gill, 2017). Actuarial analysis was emphasised and implemented as a relevant metric to measure outcomes. Actuarial analysis involves calculation of long-term projections of costs and liabilities, estimating the "risk" that an individual will make a claim or use a benefit.

Table 2.1. Timeline of the development and implementation of the New Zealand social investment approach

Year	Development
Social Investment—improving welfare and reducing fiscal liability	
2011 – 12	Introducing Social Investment for welfare
Social Investment—improving social and economic outcomes for complex families	
2015	Introducing Social Investment for housing
2015	Speech: Speech to the Treasury Guest Lecture Series on Social Investment, Bill English
2015	Analytical paper: Using Integrated Administrative Data to Identify Youth Who are at Risk of Poor Outcomes as Adults
2016	Report: Defining Social Investment Kiwi Style, New Zealand Institute for Economic Research
2016	Introducing the SIAL (Social Investment Analytical Layer)
2016	$652.1m Social Investment package for vulnerable New Zealanders
2017	Report: Governance and accountability in Social Investment, Information release, The Treasury
2017	Social Investment Agency established
2017	Speech: Changing Lives: Our Social Investment Approach, Amy Adams
Labour Government elected. Social Investment - improving wellbeing	
2017	Introducing "Investing for Social Wellbeing"
2018	Programme: "Your Voice, Your Data, Your Say" community engagement
May – September 2018	Investing for Social Wellbeing community engagement
May – September 2018	Data Protection and Use Policy community engagement
November 2018	Report: "What You Told Us" - Findings of the "Your Voice, Your Data, Your Say" engagement on social wellbeing and the protection and use of data, Social Investment Agency
February 2020	Report: Towards Better Social Sector Decision Making and Practice: A Social Wellbeing Approach
March 2020	New Zealand Social Investment Agency renamed Social Wellbeing Agency
December 2020	Social Wellbeing Agency releases the Dataset Assembly tool

Model 2, which emerged during 2015-2016, took a broader approach. While still focused on reducing fiscal liabilities, more emphasis was placed on improving social and economic outcomes for vulnerable and complex families, and generating wider social benefits. In a 2015 cabinet paper, the then Minister of Finance, Bill English and the Minister of State Services, Paula Bennett defined social investment as follows:

> *Social investment is about improving the lives of New Zealanders by applying rigorous and evidence-based investment practices to social services. It means using information and technology to identify those people for whom additional early investment will improve long term outcomes, better understanding their needs and what works for them, and then adjusting services accordingly. What is learnt through this process informs the next set of investment decisions. (Cabinet Social Policy Committee, 2015: 1)*

The new model saw a broadening of the target population to include children at risk, broadening of goals beyond long-term fiscal liability and a broadening of the range of analytical tools employed, including an increased focus on cost-benefit analysis (Boston & Gill, 2017). As a newly elected leader of the National Party in 2016, Bill English announced that, under his leadership, the government would prioritise the social investment approach (Destremau & Wilson, 2017). While the emerging second model of social investment represented a broadening of both the target population and target metrics, Boston and Gill (2017) argue that the model was still underdeveloped, lacking in specificity, and not fully embedded in the public management system.

In 2017, New Zealand voters elected a new Labour Government after 10 years of National Party rule. Whereas social investment was a high priority of the outgoing government, championed by the former Prime Minister Bill English, the incoming government, under the leadership of Prime Minister Jacinda Ardern, developed an adjusted approach referred to as “Investing for Social Wellbeing” (Cabinet Social Wellbeing Committee, 2018). The Labour Government highlighted this new approach as “a marked departure from the previous social investment approach,” noting that it moved away from “a narrower focus on reducing government costs and forward liability” to focus on wellbeing more broadly (Cabinet Social Wellbeing Committee, 2018).

The 2017 October election may be seen as marking the end of the development of Model 2, as delineated by Boston and Gill (2017), and the beginning of a new approach focusing on wellbeing rather than welfare. Whether this should be called the third phase of the NZ Social Investment approach or the end of the NZ Social Investment approach is debatable. After the election, community engagement projects with a focus on protection and use of data were initiated in the Social Investment Agency. In March 2020, the agency was renamed the Social Wellbeing Agency, marking an end to the language of “social investment” in the New Zealand Government. However, this has not meant the end of activities connected to a social investment type approach. The Social Wellbeing Agency has continued work on improving tools for the use of government data for research and insights, such as the Dataset Assembly tool released at the end of 2020. The tool is designed to standardise and automate the process of combining data from a wide range of sources into a single useable dataset (see https://vhin.co.nz/dataset-assembly-tool/).

One thread of continuity across all phases of the development and implementation of the social investment approach to welfare in New Zealand is the use of administrative data to identify groups at risk and generate evidence for policy and program outcomes. The IDI compiled by Statistics New Zealand is the key data resource developed and championed by the New Zealand Government for these purposes (Stats NZ, 2020). The IDI links individual-level data from different government agencies, including tax, health, education and social services, with data from the 2013 and 2018 censuses and other national surveys (Milne et. al., 2019). The IDI covers all New Zealand residents, is longitudinal and is regularly updated and expanded with new data sources. As might be expected, linkage rates and data quality vary for different surveys and across social groups (Milne et. al., 2019). When asked about how to create evidence-driven approaches that sustain beyond changes of government and political machinery, Bill English, one of our interview participants, replied:

> *So, here is the key thing you have to do. Well, there’s two bits that are absolutely key to sustainability. One is, institutionalise the integrated data. Because then you have a body of knowledge which they can look away from, they can steer around, but it exists. And it’s very hard for people to un-know … So you have to institutionalise it, and that has been done successfully in New Zealand. The second thing is, you have to create external demand … So, we went hard*

> *on selling it to the social science community and some of their key people, including the chief government scientist and chief social scientist, who are very well respected people.* (Bill English)

English's perception of the value and importance of the IDI was supported by comments from other interview participants. As discussed below, the IDI is widely perceived as a comprehensive source, but with some important limitations, including data quality and infrastructure, as well as lack of understanding amongst policymakers about what the data are able to say about complex social lives and outcomes.

OUR CASE STUDY APPROACH

Our aim in this chapter is to describe the social investment approach in New Zealand, with the aim of identifying strengths, weaknesses and lessons learned, particularly for the Australian policy context. We commenced with a desktop review of key policy papers, documents and literature to understand the historical and political context underlying the development and implementation of the social investment approach in New Zealand.

We then identified key stakeholders for interview in order to gather information about how the approach was received in New Zealand, some of the perceived strengths and weaknesses, and the key issues, as identified by those actively involved in developing, implementing or applying the new approach. In order to cover a wide range of opinions and viewpoints, we mapped out stakeholders who had been involved in different ways in the social investment approach, including senior political and government figures, policy analysts, academic researchers and community representatives. The identification of relevant stakeholders was done through desktop research, review of key literature and references and personal recommendations. For the purposes of this chapter, we have used alphabetical letters rather than names when referring to our interview participants, with one exception. We were fortunate to interview Bill English, former New Zealand Prime Minister and one of the architects and key proponents of the social investment approach. Given the likelihood that his comments will be easily identifiable and the importance of being able to name him directly as one of our participants, we sought and received his permission to report his involvement in the study and identify any comments from his interview quoted in this chapter.

Each interview participant was emailed and provided with an outline of the research and an interview consent form. The interviews were undertaken over the phone or by Skype/Zoom in April/May 2019 and were recorded and subsequently transcribed. We identified five selected participants, all of whom accepted our invitation to be interviewed for the study. All of the participants had been involved in the social investment approach in one form or another, including its development, implementation, or through research using the IDI. In several cases, our participants had been involved in more than one of these areas. Each interview took approximately one hour and included discussion of their understandings of how the New Zealand social investment approach had developed, their perceptions of its strengths and weaknesses, their perceptions of the implementation process and modifications to the approach over time, and their knowledge and views about the quality, accessibility and importance of the data infrastructure underlying the approach. Each interview covered slightly different issues, according to the context of the participant and their involvement with the social investment approach. A draft of the chapter was sent to each participant to provide an opportunity for them to comment on our interpretation of the data and to ensure their agreement with our representation of their views. No substantive changes were requested by any of the participants.

THEMES AND KEY FINDINGS

Defining the New Zealand social investment approach

In a 2015, a New Zealand Cabinet paper headed by the then Minister of Finance, Bill English and the Minister of State Services, Paula Bennett described the benefits of social investment in the following way: (1) The social investment approach sets clear and measurable goals. (2) It provides a better use of information and technology to understand the needs of people who rely on social services. (3) It provides systematic measurement of the effectiveness of services, and feeds those learnings back into the decision-making process. (4) Social investment enables governments to purchase outcomes rather than specific inputs, shifting funding to the most efficient service providers, whether government or non-government providers (Destremau & Wilson, 2017). These sentiments were echoed in our interview when we asked Bill English to describe the New Zealand social investment approach.

> *It gives you a framework which shows you who to apply it [a policy programme] to and then it gives you a feedback loop that makes sure you stick at it and that you adjust policy as time goes on. Now, that's the heart of the implementation of social investment … It does not, in itself, provide a new service, but rather, shows you investment and pay-offs for people in existing services. This way, social investment is meant to identify services which are not capturing the whole population in need, so those most in need miss out, and likewise, when services are capturing people who would otherwise get on fine. (Bill English)*

Compared to Bill English, other participants were a bit more cautious about the benefits of the social investment approach. Our researcher and community representative suggested:

> *The social investment model was a way of -- a benign description of it would be: informing early investment to pre-empt later costs to the state. And a less benign description of it would be identifying high-risk individuals who are going to be really, really expensive. (Participant A)*

Another questioned whether social investment was actually a new approach or simply a repackaging of previous approaches. Participant C, a university-based researcher with previous experience working in government agencies, was quite adamant that social investment for welfare was not new.

> *There is nothing inherently ground-breaking here at all. The entire notion of a social investment is overblown. (Participant C)*

And later continued:

> *I am not a fan of the idea that there is a social investment approach or in any way represents some revolutionary approach in social policy. I think it's very sensible to use more data and empirical evidence in policy, and there's some sensible questions there. But I think the entire notion of a social investment approach is overblown … I'm quite sceptical of the idea of the social investment approach as a thing… The idea of a social investment approach as a particular way or methodology … like with some solution there, is nonsense. There are some good ideas there, there are some things that are worth lifting out, but it's not a transformation on social policy. (Participant C)*

Similar views that social investment was not a particularly new approach are also apparent in some of the literature. For example, Destremau and Wilson (2017) note that the idea that some social spending can be considered social investment rather than social cost can be traced back to the writings of Swedish economist Gunnar Myrdal in the 1930s.

Debate about the definition of the social investment approach seems, to a large extent, focused on whether it should be seen as a purely fiscal approach to reduce government spending or something broader, aimed at improving lives. For example, one of our interview participants suggested that social investment is worth doing because it helps develop long-term thinking about fiscal spending, but only provided that you recognise it for what it is, a better measure of fiscal cost.

> *One thing [that can lift the quality of policy advice] is getting a better handle on fiscal impact, particularly fiscal impact over time, and recognising that spending now … on someone with a service is not a good measure of lifetime or ongoing fiscal spend … That is pretty close to the original Bill English-era idea of social investment. And I think that's worth doing, provided you recognise what it is. It's giving you a better measure of a fiscal cost. (Participant C)*

Another suggested that disagreement about the definition of social investment is not just a political issue, but also has consequences for how it is implemented, resourced and managed. Lack of a shared understanding about the aims may lead to variations in implementation and levels of support across agencies. For example, one of our participants with experience in government, non-government organisations, as well as business, argued that the way in which social investment was applied in some government agencies, such as the Ministry for Social Development, was indeed just a fiscal approach. But in other areas, it was not.

> *They're [social investment and a fiscal approach] not the same thing. The way social investment was being applied in the justice sector and in the child protection sector wasn't a fiscal approach … Rather than the single fiscal objective, they had a range of objectives that they were pursuing, and they were measuring each of them. (Participant D)*

This participant also argued that criticisms of how social investment was implemented in certain areas were not necessarily applicable to how it was implemented in other areas, where they were not using purely fiscal objectives. Undoubtedly, there was a wide range of views and understandings about the social investment approach at all times, and the aforementioned distinction between Model 1 and Model 2 implies a clear temporal distinction that, in reality, was possibly much less defined. Both views may have coincided across different sectors and within sectors at the same time. Some people may have viewed social investment as a fiscal policy, while others viewed it as a means of supporting social and economic improvements for vulnerable and complex families. Criticism of the social investment approach to some extent then depends on what "phase" is being assessed: the "when" and "who" is making the argument.

In light of these disagreements about what social investment actually is or was, and, further, the trouble in agreeing on what is meant by target measures, one participant noted a lack of policy evaluation of social investment.

> *[There is] [n]o actual analysis deconstructing what "core concept" means. What do you mean when you say "return"? Do you mean fiscal return or broader social return? Discussions about the conditions under which fiscal return would be a good proxy for social return. When does early intervention make sense? (Participant D)*

Another of our participants, head of a social research agency at a university, emphasised the actuarial basis of the approach, suggesting that the core issue was not about the value of social investment for reducing state fiscal cost, but rather, ideological differences in ways of thinking about risk and redistribution.

> *The focus on cost, I don't think, was the thing. The focus was on this kind of actuarial or insurance-based approach. There's two ways you can think of our safety nets. You can think of our safety nets as a system of redistribution, which is a very static concept, like there are some people doing well, some people are not doing very well. Let's take from the people doing well and make sure that the people who are doing badly are not suffering from whatever is happening in their lives. Or you can take it as an actuarial model, which is: some people's house didn't burn down this year, some people's house did burn down. What we're doing is insuring the people whose house burned down .... It was a very individualised insurance model. (Participant E)*

The implementation of actuarial analysis, according to Bill English, came from the use of this type of actuarial framework in the Accident Compensation Corporation (ACC) scheme, a national injury compensation scheme primarily funded through levies, designed to compensate residents that have suffered personal injuries (Accident Compensation Corporation, 2020). However, the social investment actuarial approach was supposed to be different from the ACC and other general insurances that do top-down risk analysis. Instead, the idea was to focus on building up a picture of big risks from individual trajectories and to understand cumulative effects on particular populations. The implementation of actuarial analysis as the main mode of analysis within the social investment approach met substantial scepticism, while others were merely cautious (Boston & Gill, 2017). Our interview participants also expressed a range of views. One participant expressed scepticism, also picking up on the ideological component of this divide.

> *I'm pretty sceptical of the whole initial thing of social investment as actuaries ... But I'm also not entirely sceptical that it's worth looking at forward fiscal costs associated with programs. It's an area which has divided people quite strongly into different camps. I mean, economists have generally been pretty sceptical of it. Possibly a little bit too much so. But, then again, on the actuarial side, I think they've been vastly too ideological and over-confident with it. (Participant C)*

Bill English noted that implementing the social investment approach shone a light on the difficulty in defining and measuring concepts that at first glance seem straightforward.

> *And it turned out, for instance, in that case, teachers, police, and nurses all had quite different tools in their head for determining whether a child was an abuse victim. And that took 12 months to sort out. But it wouldn't have come up if we didn't start the conversation. ... They're all talking about child abuse and we all thought they knew and we knew what we were talking about. But it turned out, we didn't know. (Bill English)*

In some of our interviews, issues of transparency were also raised. One participant felt that the social investment approach had been characterised by a lack of transparency in the methods being used.

> *One concern was, there was very little transparency about their work. What does their algorithm look like? What does the code look like? From what I could see, there was very little engagement with academia and very little engagement with external peer review. (Participant A)*

The issue of transparency is important for several reasons, including its effect on the legitimacy of ensuing policies or approaches. As we have seen, there does not seem to be agreement on how data should be analysed, given the criticism of the actuarial approach. Methods of analysis for the Integrated Data Infrastructure will depend partly on your standpoint. For example, should welfare systems be treated more like market insurance schemes or redistribution systems, and how do you define and measure outcomes? There are many small decisions in adjusting and specifying a model throughout the process of data analysis that will have important consequences for what your model does and how to interpret it. When these decisions are not transparent from the outside, peer review and critical assessment of the work becomes impossible.

> *Working with actuarial risk-profiling models means even less transparency because the financial services firms can regard their work as proprietary. (Participant A)*

This point has become especially important for lessons learned for the future, in particular, where lack of transparency relates to proprietary code.

In sum, there is not, and maybe we should not expect, consensus on what a social investment approach is, how "new" it is, whether it is a fiscal policy designed to reduce costs, whether it is aimed at social and economic wellbeing, and whether the actuarial approach is a suitable one. What we can learn is that all analytical decisions, including implementation processes, will be based on core assumptions and ideological beliefs about how welfare systems should be managed, what success looks like and how it will be achieved or measured. If there is a lack of clear shared vision and understanding, implementation processes will be compromised. Clear, visible, transparent and shared understandings are important to enable constructive opportunities for critical peer review, informed discussion and new approaches going forward.

Implementation and the New Zealand Social Investment Agency

An important part of the way the New Zealand Government decided to manage implementation of the social investment approach was to move its base out of the Ministry for Social Development and establish a standalone agency, called the Social Investment Agency. The new agency, established in 2017, was tasked with furthering the implementation of the approach and was intended to broaden the use of social investment across agencies. The Social Investment Agency continued under the Labour Government, with the stated aim of "strengthening the use of data, analytics and insights in social sector decision making to improve New Zealanders' wellbeing" (Social Investment Agency, 2019), but was renamed the Social Wellbeing Agency in March 2020 (The Beehive, 2020).

While the goal of the Social Investment Agency was to drive the implementation of the social investment approach, the view amongst our interview participants was that this decision severely weakened the possibility of achieving effective implementation. One participant commented, "Original name, same old people" (Participant A). Another stated, "There has been a lot of turnover at the top-tier staff since it started up. That doesn't allow you to build much capacity" (Participant C). And another: "It reads to me like something that is slowly deflating" (Participant D). One issue for the agency was that it was set up quite late in the social investment timeline, leaving it little time to get established and running before the change of government in 2017 and the ensuing change from Social Investment for Welfare to Investing for Social Wellbeing. According to some of our interview participants, the Social Investment Agency did not manage to establish itself as the place to go to achieve impact. Its staff were not perceived as the "people who can move the dial" (Participant E).

> *I think there was a sort of sidelining of the SIA and it didn't have the intellectual heft to be able to go and impose itself on big operational behemoths, such as the welfare system. It was underpowered for what it was trying to do. (Participant E)*

In a similar vein, Participant D commented:

> *The problem that agency will have is it's kind of a second-tier agency. So, they're not at the big table … the SIA was a small and vulnerable agency. They did not pick fights with the big players in town, because they could just be shut down. (Participant D)*

Participant D emphasised the fact that to achieve systematic policy change, you need both strong political leadership and strong bureaucratic leadership pushing for the same change.

> *If you don't have a capable bureaucratic leader, you don't get effective change. Creating a bespoke agency like the Social Investment Agency is the wrong solution to that, because they are small and vulnerable. … If you are a small agency, you choose your fights very carefully. (Participant D)*

Additionally, our participants perceived that there were few incentives for cross-agency collaboration, and that agencies perceived sharing data and insights as a potential risk to their autonomy and future funding. Further, if data was shared in an incorrect way, there was the possibility of media scandals. Instead, it was easier to avoid these risks by not sharing data and not working across agencies. Bill English noted these power struggles, which he suggested originated out of competition for funding, acting as roadblocks in implementing insights, particularly in cases that involve many agencies, such as cases of family violence or child abuse, where a number of different agencies were involved in the response.

> *You can say, "Okay, this is what we now know about the bottom 1% of kids." Then it's just 2 years of chaos while agencies fight. So, the problem's here a fight over why you shouldn't unbundle their funding and who's going to control it. (Bill English)*

This suggests that incentives for collaboration are needed, and that, while cross-agency solutions come with difficulties, leadership, clear processes and shared understanding of goals are essential. What is also clear from this section is the importance of establishing the right governance and bureaucratic structures. The danger of developing a standalone agency that can be easily sidelined is apparent from the above comments, and points to the need to ensure that the department tasked with implementing an approach such as social investment must be located in a core, whole-of-government agency and funded appropriately to do the task.

Data infrastructure and competency

A key component of the social investment approach in New Zealand is the Integrated Data Infrastructure. This is a population-wide person-centred micro-dataset linking data from a range of government agencies, surveys from Statistics New Zealand, including the 2013 and 2018 censuses, and data from non-government organisations (Stats NZ, 2017). Funded by the New Zealand Government, the IDI was promoted as being developed with the aim of helping decision makers find answers to complex social questions (Stats NZ, 2017). While the IDI was established and is run independently of the Social Investment Agency, it was fundamentally a part of the social investment approach to integration of data into policy, and achieving the key goal of identifying and tracking population cohorts over time and the opportunity to evaluate interventions or system-wide policy changes (Milne et al., 2019). It continues to be the centre of attention under the Social Wellbeing Agency, with ongoing release of improvements and tools. The social investment approach established a political desire for integrated data, which, in turn, has had positive outcomes for the sustainability of the IDI.

> *That, I think, was a really amazing legacy from the National Government and from Bill English, that desire to have this integrated data, and the social licence that they got around it from the agencies who are willing to put the data in. (Participant E)*

The IDI was unanimously perceived as a great resource by all of our interview participants, illustrating that a government-led initiative has potential to lead to better collaboration and access. One participant stated that Australians should "bully, bribe, beg" to get the Australian Bureau of Statistics to put social surveys together with linked administrative data (Participant C). We note that considerable progress in this direction has occurred in Australia through the Multi-Agency Data Integration Project (MADIP), which includes data on health, education, income support payments, employment and demographics (ABS, 2021). Another participant (Participant E) highlighted the team at Statistics New

Zealand for their work showing that backend integration of administrative and survey data does not have to be difficult and is achievable. The IDI was considered an outstanding resource both for government and researchers. As noted, Bill English specifically described the IDI as a strategy for making the social investment approach possible and for helping to ensure the longevity of the approach beyond changes of government.

> *If the government tried to dismantle the IDI, there would be outrage … So that's why you've got to create external constituencies for the information. (Bill English)*

Nevertheless, the IDI has experienced issues with infrastructure, with several of our participants pointing out that it "drops out every few weeks" (Participant C) or "collapses once a fortnight" (Participant A). This recurring instability of infrastructure capacity was perceived as a result of a lack of funding of Statistics New Zealand, with insufficient resources put towards the infrastructure underlying political initiatives such as the social investment approach. As one participant put it, "Without good quality data, all this complicated complex modelling that we're doing is pretty worthless" (Participant A). Funding cuts and changes to how Statistics New Zealand collected census data in 2018 were perceived as particularly problematic for the Indigenous population, since it affected Indigenous data coverage to a much greater extent than the rest of the population.

A further perceived issue with the implementation of the social investment approach, and the establishment of the Social Investment Agency in particular, was a lack of competency in data analysis and interpretation, and limited understanding of the possibilities and limitations of administrative data.

> *I just feel that they didn't have the technical expertise. There was lots of people who didn't seem to have a lot of data knowledge, really. (Participant E)*
>
> *It hasn't actually yet built a culture of using data. (Participant C)*
>
> *We need more people, decision makers inside government and outside government, with statistical literacy and an understanding of what big data can do, and, more importantly, what it can't do. (Participant A)*

The lack of data and analytical competency in the Social Investment Agency and the lack of broad leadership were seen as major roadblocks for the success of the social investment approach, making it hard to create a culture of data and evidence-based policy. Participant C emphasised the "lack of understanding of quantitative methods" and the value of using them from the "upper levels of civil service" as the main barrier in the continued development and implementation of the social investment approach.

Moreover, our participants noted that even if the IDI is a great resource for policy development, the issue of competency and understanding of big data is integral for its incorporation into social policy.

> *It's like any tool. We're learning its role and we're learning more about its limits. Is it a positive development for policy? Absolutely. Have we achieved its full potential? No. Is it the greatest thing since sliced bread? No. (Participant D)*

The researchers amongst our participant group were very clear that it was important to understand the datasets, how they were constructed, what the variables mean, and what's not there. Skilled users with an understanding of these datasets can pull a lot of evidence out of the IDI, but limitations in analytical skills and lack of understanding of the data were viewed as major limitations.

> *So we're running these highly complex models that are calculating these fabulous betas and particular point estimates, but we don't know what the confidence intervals are. (Participant A)*
>
> *The fundamental problem is going from correlation to causation. There is nothing inherent in the IDI that answers that question. There is nothing in the big datasets telling you what to do about something. (Participant D)*

Overall, the three key messages from our discussion of the Integrated Data Infrastructure were: first, that it was a wonderful resource with a great deal of potential for supporting social policy analysis generally, including social investment; second, that effective use of the data required a combination of specific technical skill sets, a broader culture of evidence-based policy integrated at all levels, and competency in interpreting and translating research into policy; and third, an effective social investment approach required proper funding for the underlying data infrastructure, including strong data governance and ongoing investment that maintained the data, kept it up to date, and supported effective usage in terms of training, information and culture.

Social licence and community data rights

An important issue in the use of big data for social policy is social licence to use the data, particularly licence from vulnerable groups and communities who are highly represented in the data. Data ownership has emerged as a major concern for the UN Permanent Forum on Indigenous Issues (UNPFII) (Davis, 2016). One of our participants who was Māori raised these issues in a number of places in the interview. This participant took particular issue with the ease in which the government was willing to share the New Zealand data with international researchers, in light of the Indigenous struggle for their right to control their data.

> *Indigenous people have been asking for access to this information for years and never got it. They're doing work on our vulnerable children and giving overseas access to information about our children.* (Participant A)

He explained that the tribal nations and the iwi leaders' forum in New Zealand now consider data to be so fundamental to Indigenous development that they have created a sub-group to work with data and data sovereignty. In the 2015 charter establishing Te Mana Raraunga—the Māori Data Sovereignty Network—data is determined to be living tāonga (Kukutai & Taylor, 2016), meaning something that is considered to be of social or cultural importance (Moorfield, 2021). The Australian equivalent is the Maiam nayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Collective formed in early 2017 to develop Aboriginal and Torres Strait Islander data sovereignty principles and to identify Aboriginal and Torres Strait Islander strategic data assets (see https://www.maiamnayriwingara.org/). Kukutai and Taylor (2016) write that while Indigenous peoples have long claimed sovereign status over their land and territories, missing from the conversation has been the rights of Indigenous peoples relating to the collection, ownership and application of data about their people. Our Māori participant was critical of the way that Indigenous data sovereignty had been handled in the social investment approach, referring to it as "very disempowering for Māori," especially in relation to analyses of vulnerable children.

> *… it led to some high-level discussion about who can tell our stories with our data. We've been clear with the Office of National Statistics that no regional stories can be told about children without the engagement of the tribal nations.* (Participant A)

There were also fears that the use of predictive analytical models leads to increased stigma rather than improvements in living conditions. A key component of the social investment approach is the ability to segment populations. But this has the risk of individuals being classified as risky and being further stigmatised, without engaging with these so-called "risky groups or communities" to hear their perspectives and to work together to improve outcomes.

Our Māori participant perceived this failure to obtain social licence, from both Indigenous peoples and other populations, to be a key lesson to be learned from the New Zealand example of implementing a social investment approach.

> *The main lesson is that the people who are providing the information and the people about whom the study is, or the subjects of the study, needs to be engaged with, and not persuaded, but engaged with and supportive of both the outcomes and the algorithms being used to examine those outcomes … Doing it in a detached way will lead to incompetence at best and community disempowerment at worst.* (Participant A)

Some work on strengthening the social licence dimension of the usage of administrative data has been done by the New Zealand Labour Government as a part of their efforts in shifting from Social Investment to the Investing for Social Wellbeing approach. The "Your Voice, Your Data, Your Say" program was undertaken during 2018 as part of a process to create a data protection and use policy (Social Investment Agency, 2018). One aim of the policy was to highlight Māori and other voices and perspectives. Among the highlighted outcomes of this social licence program was the expressed need for more policy solutions that were culturally sensitive and locally led, and both the Māori groups and

Pacific peoples groups emphasised issues of trust in government and the need for culturally sensitive measurements. The report explicitly argued that these culturally sensitive measures do not only need to take the Māori and Pacific peoples' needs and wants into account, but must be developed in conjunction with representatives of these groups as true partnerships (Social Investment Agency, 2018).

Analysis without engagement can be disempowering, especially for vulnerable populations. Building trust and partnerships is imperative for establishing a social licence to use data. Countries with Indigenous peoples have an additional dimension of Indigenous rights that needs to be taken into account (United Nations, 2007). The right of Indigenous peoples to define their own measures of success have a potentially critical impact on how success measures are created and employed in big data for social policy approaches. This dimension of the social investment approach has been surprisingly invisible in government reporting and academic debate around the outcomes and success of the social investment approach and has important implications for the Australian case. We note, however, that the Indigenous Data Sovereignty discussion is gaining greater prominence in Australia, with the Academy of Social Sciences partnering with key Indigenous leaders and national data agencies to host a roundtable in mid-2021 to develop core principles for Indigenous data infrastructure.

FINAL THOUGHTS

The New Zealand social investment approach provides a very useful case study for highlighting the role of administrative data for developing and informing social policy. Developed and initially implemented in 2011–12, there is sufficient time to review the successes and challenges, how the approach fared following a change of government in 2017, and development of a new agency and implementation strategies. The New Zealand case study is particularly instructive, given the parallels with the Australian Priority Investment Approach announced in the 2015-16 Federal Budget. In this chapter, we have reported on findings from a desktop review of key literature and government reports as well as interviews with five participants representing a range of views and sectors including government, the academy, social research organisations and public and private agencies. Although clearly not representative, and not intended to be, our participants provided key insights into the New Zealand social investment approach and gave us an opportunity to identify some key themes about the use of big data for social policy. In this final section, we highlight three overarching themes emerging from the issues raised above.

First, our analyses highlight the importance of shared understandings, definitions, vision and understanding of how data and evidence can be used to inform a social investment approach. This is critical for ensuring that the approach and infrastructure underneath it is implemented appropriately, with shared understandings and sufficient resources to maintain them over time. This point is illustrated by Bill English above, when referring to defining child abuse and the importance of starting the conversation which led to identification of key differences in views across sectors. Clarity in core concepts is also critical to ensure the possibility of constructive criticism and critique of the approach from the surrounding expert community, which is integral in both developing and improving the approach, but also from a democratic transparency perspective. This is not an easy task, as new approaches are rarely developed overnight but evolve and develop over time as they are implemented, reviewed, critiqued, developed further, tested and redesigned. At any one point in time, definitions will be undergoing change and development and stakeholders will view the same approach from widely varying perspectives. Opportunities for open debate, constructive criticism and ongoing strategies to minimise division and dismissal of new ideas are essential.

Variations in perspectives were apparent throughout our case study. This emerged clearly in key texts such as the volume by Boston and Gill (2017), as well as amongst our interview participants. Some interview participants were more critical than others, and even when positive in their comments, were clearly prioritising different aspects of the social investment approach and exhibiting variations in their understanding of it. All were positive about the Integrated Data Infrastructure, but there were concerns about whether the data infrastructure was sufficiently resourced and maintained. Similar discussions have been surfacing in Australia, where the former national statistician David Kalisch publicly criticised funding cuts to the Australian Bureau of Statistics, saying that they threatened key data releases (Wright, 2019). While our stakeholders generally agreed on the importance of a stable and properly funded data infrastructure for an approach such as social investment to yield any results, we observed differences in views about the value of the data in terms of the lack of transparency in background material to ensure accurate interpretation of results and translation of results into policy outcomes, variations in views about data competency and skills, and variations in concerns about social licence and ethics of data use. In our view, these concerns do not undermine the importance of a data infrastructure such as the IDI, but rather, point to issues that must be addressed to ensure appropriate usage, governance and maintenance.

Second, governance and infrastructure are critical elements in the successful delivery of social policy approaches, and this was particularly apparent for the New Zealand social investment approach. Despite strong championing of the approach from Bill English and other senior political and government leaders, the infrastructure for building and maintaining the Integrated Data Infrastructure was viewed by our participants as insufficient, with too little funding provided to Statistics New Zealand to ensure a stable, continuously funded and well-maintained data infrastructure. Moreover, participants were clear that the movement of the Social Investment Agency to a standalone agency had severely compromised its power, resources and ability to deliver on the policy approach. We will not be the first to highlight the importance of governance arrangements in delivering policy, but we see this as a very clear example of how governance decisions can effectively undermine, or, at the very least, sideline previously robust social policies.

Third, underlying all social policy approaches, of course, are political frameworks, priorities, opportunities and constraints. Bill English, both as National Party Finance Minister and later as Prime Minister, was a key champion of the social investment approach and was, and is, passionate about its promise to find solutions to some of New Zealand's most complex and wicked social problems. While aspects of the social investment approach survived the change of government in 2017 and the election of Prime Minister Jacinda Ardern and a Labour Government, the approach has shifted markedly to a focus on wellbeing. More than a semantic shift, this change has coincided with a downplaying of the social investment approach to the point where some of our interview participants commented on its overall deflation and dwindling. Clearly, the New Zealand social investment approach is evolving over time and will continue to do so in relation to changing political and economic contexts. The IDI, on the other hand, is potentially an enduring resource that is continually developed under the Social Wellbeing Agency, with much broader value than application to the social investment approach. There will inevitably be political debate about the priorities between competing policy concerns, but all depend upon the IDI, or something similar, and on appropriate management of the related ethical issues.

Fourth, although evidence can point to solutions that may improve individual outcomes, translating this evidence into actionable policy is not straightforward. As Gluckman (2017: 6) notes, there are many complexities for effective social policy including "how to juggle constituent demands, partisan ideology and goals, and real world fiscal and societal constraints." Bill English makes a similar point:

> *The way to think about the social investment framework is that it basically says, "Find a solution that works"' ... the hardest job is not actually solution finding. Those emerge. It's changing existing institutions to accommodate those solutions. (Bill English)*

While big data may offer increasingly fine-grained insights into what works, for whom, when and where, translating these insights into real world solutions will rarely be straightforward.

ACKNOWLEDGEMENTS

We would like to thank the Australian Research Council Learned Academies Special Projects (LA170100011), the ARC Centre of Excellence for Children and Families over the Life Course (CE140100027) and the Academy of Social Sciences for their support of this work. We are grateful to our colleagues on the Learned Academies Special Projects grant: Professors Janet Chan, Peter Saunders, David Abramson, Ross Homel, Sheila Shaver and Mark Western, their research teams and the project members of our Advisory Group for their collaboration, comments and support. In particular, we thank Dennis Trewin and Andrew Podger for their comments on an earlier draft. Most of all, we thank our interview participants for the generous and timely way in which they engaged with this project and for sharing their views and thoughts in an open and thoughtful manner. Any errors or inaccuracies are ours alone.

REFERENCES

ABS (2021). Multi-agency data integration project (MADIP). https://www.abs.gov.au/websitedbs/d3310114.nsf/home/statistical+data+integration+-+madip+research+projects

Accident Compensation Corporation (2020). What we do. https://www.acc.co.nz/about-us/who-we-are/what-we-do/

The Beehive (2020). Press release: Social Wellbeing Agency replaces Social Investment Agency with new approach. http://www.beehive.govt.nz/release/social-wellbeing-agency-replaces-social-investment-agency-new-approach

Boston, J., & Gill, D. (2017). Social investment: A New Zealand policy experiment. Wellington, NZ: Bridget Williams Books.

Cabinet Social Policy Committee (2015). Implementing social investment. Wellington, NZ.

Cabinet Social Wellbeing Committee (2018). Toward investing for social wellbeing. Wellington, NZ.

Destremau, K., & Wilson, P. (2017). Defining social investment, Kiwi style. In J. Boston & D. Gill (Eds.), Social investment: A New Zealand policy experiment (pp. 35-74). Wellington, NZ: Bridget Williams Books.

Davis, M. (2016). Data and the United Nations Declaration on the Rights of Indigenous Peoples. In T. Kukutai & J. Taylor (Eds.), Indigenous data sovereignty: Toward an agenda (pp. 25-38). ANU Press. http://www.jstor.org/stable/j.ctt1q1crgf.9

Elder, G. H. (1994). Time, human agency, and social change: Perspectives on the life course. Social Psychology Quarterly, 57(1), 4-15. https://doi.org/10.2307/2786971

Heckman, J., & Cunha, F. (2010). Investing in our young people. In A. J. Reynolds, A. Rolnick, M. Englund, & J. A. Temple (Eds.), Cost-effective programs in children's first decade: A human capital integration (pp. 381-414). Cambridge, New York: Cambridge University Press.

Gluckman, P. (2017). Using evidence to inform social policy: The role of citizen-based analytics. A Discussion Paper. Office of the Prime Minister's Chief Science Advisor. www.pmcsa.org.nz

Kukutai, T., & Taylor, J. (Eds.) (2016). Indigenous data sovereignty: Toward an agenda. Acton, ACT: ANU Press.

Mayer, K. U. (2004). Whose lives? How history, societies, and institutions define and shape life courses. Research in Human Development, 1(3), 161-187. https://doi.org/10.1207/s15427617rhd0103_3

Midgley, J., Dahl, E., & Wright, A. C. (2017). Social investment and social welfare: International and critical perspectives. Cheltenham, UK: Edward Elgar Publishing.

Milne, B. J., Atkinson, J., Blakely, T., Day, H., Douwes, J., Gibb, S., . . . Teng, A. (2019). Data resource profile: The New Zealand integrated data infrastructure (IDI). International Journal of Epidemiology, 48(3), 677-677e.

Moorfield, J. C. (2021). Te Aka online Māori dictionary. https://maoridictionary.co.nz/search?idiom=&phrase=&proverb=&loan=&histLoanWords=&keywords=t%C4%81onga

Nolan, B. (2013). What use is 'social investment'? Journal of European Social Policy, 23(5), 459-468. https://doi.org/10.1177/0958928713499177

Sen A. (2003). Development as capability expansion. In S. Fukuda-Parr, & A. K. Shiva Kumar (Eds.), Readings in human development.. (pp. 3-16). New Delhi and New York: Oxford University Press.

Social Investment Agency (2018). What you told us: Findings of the "Your voice, your data, your say" engagement on social wellbeing and the protection and use of data. Wellington, New Zealand. https://natlib-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=NLNZ_ALMA11319053770002836&context=L&vid=NLNZ&search_scope=NLNZ&tab=catalogue&lang=en_US

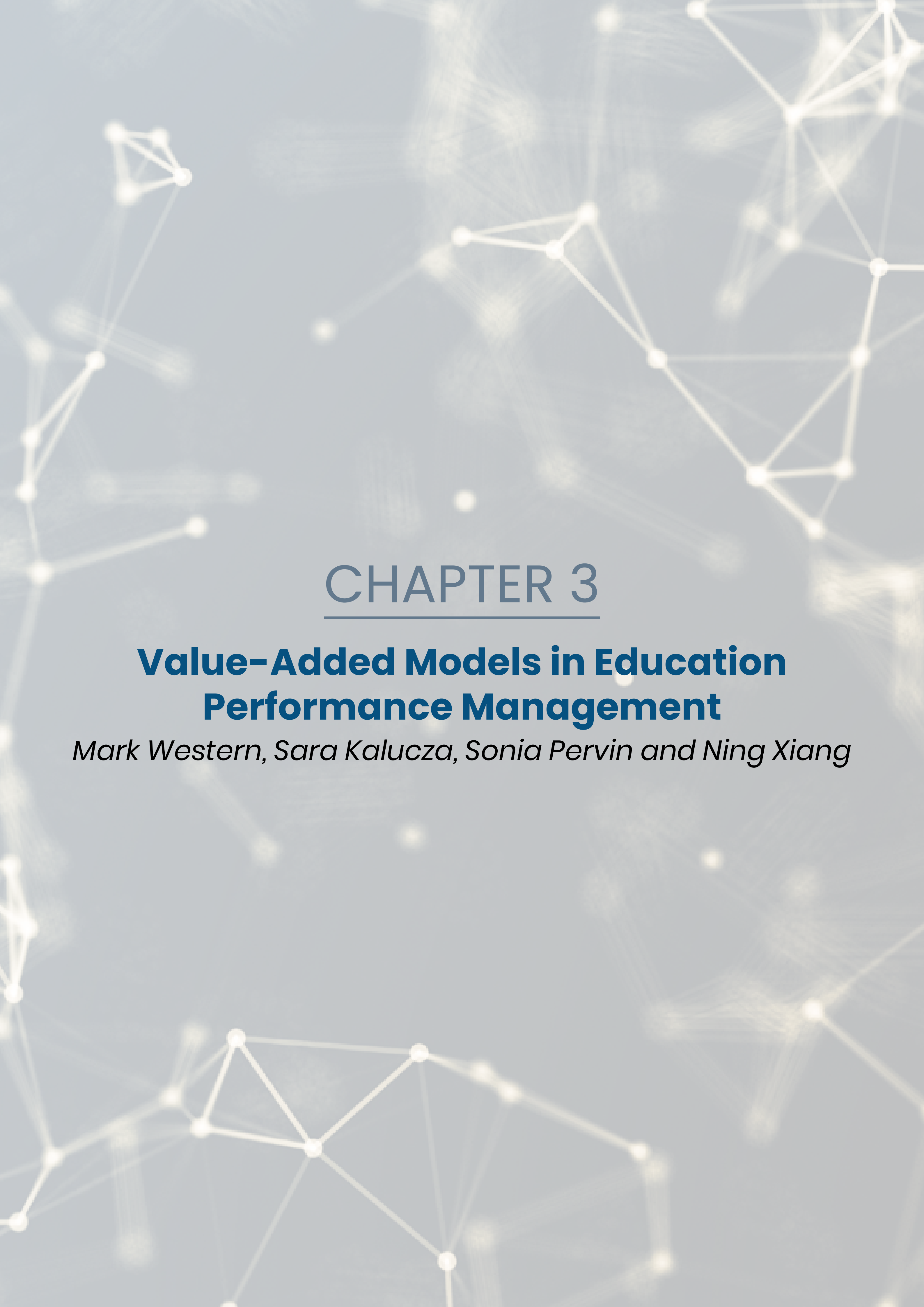
Staines, Z., Moore, C., Marston, G., & Humpage, L. (2020). Big data and poverty governance under Australia and Aotearoa/New Zealand's "Social Investment" policies. Australian Journal of Social Issues. https://doi.org/10.1002/ajs4.129

Stats NZ (2017). Integrated data infrastructure: Overarching privacy impact assessment. https://www.stats.govt.nz/assets/Uploads/Integrated-data-infrastructure/idi-overarching-pia.pdf

Stats NZ (2020). Integrated data infrastructure. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/

United Nations (2007). 61/295. United Nations Declaration on the Rights of Indigenous Peoples. https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_E_web.pdf

Wright, S. (2019). ABS boss slams a decade of funding cuts, warns key data is at risk. The Sydney Morning Herald. https://www.smh.com.au/politics/federal/abs-boss-slams-a-decade-of-funding-cuts-warns-key-data-is-at-risk-20191015-p530xw.html



CHAPTER 3

Value-Added Models in Education Performance Management

Mark Western, Sara Kalucza, Sonia Pervin and Ning Xiang

Value-Added Models in Education Performance Management

Since the early 2000s, many countries have introduced systems to measure student achievement to assess how effectively education systems improve student learning. Such systems rely heavily on data about student characteristics, teachers, schools, and student learning performance. International assessments such as the OECD's Program of International Student Assessment (PISA) (OECD, 2008), and the Trends in International Mathematics and Science Study (TIMSS) (Australian Council for Educational Research, 2020) and Progress in International Reading Literacy Study (PIRLS) (ABS, 2012), sit alongside national initiatives. The World Bank concisely summarises why such activities matter. "Measurement of education system performance is a cornerstone of improving learning. Accurate data underpins successful policies" (The World Bank, 2018).

Australia, like many other jurisdictions, links educational performance assessment to improved student learning through the mechanism of school and teacher accountability (Everson, 2017). Because teachers and schools are so important to student learning, education systems need to incorporate ways to hold teachers and schools accountable for students' learning outcomes (Kleinhenz & Ingvarson, 2004). Educational performance assessment provides one such vehicle. Australia introduced performance-based accountability policies and frameworks linked to national data and testing in 2008, via the Australian National Curriculum, the National Assessment Program, including the National Assessment Program of Literacy and Numeracy (NAPLAN), and the My School website, which allowed "like" or similar schools to be compared on NAPLAN performance (Rowe, 2017). The Australian Curriculum and Reporting Authority has governance and oversight of NAPLAN, the National Curriculum, and My School. Under National Agreements between the states/territories and the Australian Government, state and territory departments of education receive reward performance payments from the Australian Government for achieving agreed literacy and numeracy targets (Australian National Audit Office (ANAO), 2012).

Other countries introduced school and teacher accountability systems of education performance assessment earlier than Australia. The United States is one such case. There, test-based accountability systems linked to system-wide data, and so-called "Value-Added Models" (VAMs) of student learning were used to make high-stakes decisions about matters such as school-district funding and teacher pay and employment. But over several years, US policymakers realised that using Value-Added Models in this way did not improve student learning outcomes, and, in fact, had some adverse consequences for many teachers and students.

Since 2016, many US states have moved away from a primary focus on VAMs and system-wide standardised testing to assess teacher effectiveness and education system performance. In this chapter, we consider VAMs, and particularly their implementation in the United States, as a cautionary tale about the use of big data for social policy. While the use of VAMs led to a great deal of controversy, and policy and academic debate about the merits of such methods for educational performance assessment, we argue that much of this debate was misplaced because it failed to properly identify one critical issue: using data, models and algorithms for operational decision-making in public policy is a distinctive application or use case that requires particular frameworks and safeguards. Using models and data for operational decision-making about individuals, such as whether a teacher should be re-employed, is not the same as using models and data in research to inform higher-level policy making. Using data and algorithms to inform operational decisions about individuals who are subject to those decisions needs accompanying ethical and governance frameworks and standards of evidence to support the use of such approaches. After describing the VAMs case study and its implications for data and algorithm-informed decision making in public policy, we conclude the chapter by describing how health research might offer some useful correctives to some of the concerns raised by this case study.

WHAT ARE VALUE-ADDED MODELS (VAMS)?

Value-Added Models are statistical models to analyse educational outcomes like student performance on a standardised educational test. VAMs attempt to measure the impact of teaching based on students' test achievement against some estimate of how students are expected to do, considering their earlier achievement level, and, depending on the specific model, other background factors assumed to affect student learning beyond the control of teachers and schools (Brown et al., 2016). Educational decision makers use VAMs in education performance assessment because researchers and others have argued that these models validly and reliably measure school and teacher effectiveness (SAS Institute, n.d.; Wright, 2017) in contributing to student learning outcomes.

Many different VAMs are possible, depending on what variables and data are included and the kinds of mathematical functional relationships between variables that the model incorporates. However, VAMs commonly assume that the model's estimates of teacher quality or teacher effectiveness indicate how much an individual teacher contributes to (or detracts from), the gains a student makes in student learning, typically measured through standardised test scores, in some relevant population such as a state, school sector or education system (Everson et al., 2013).

Value-Added Models usually control for students' prior achievement (test scores), and sometimes, but not always, student-level socio-demographic characteristics (e.g., race, gender, ethnicity, levels of poverty, students' levels of English language proficiency, special education status). Models sometimes also include additional student-level variables such as attendance, suspension, retention records, and classroom, along with school-level variables. The value-added estimate is the teacher fixed effect or the residual that remains when all possible controls are included (Pivovarova et al., 2016).

$$A_{i\,jt} = \beta_0 A_{i,\,t-1} + \beta_1 X_{it} + \beta_2 Z_{jt} + \mu_{i\,jt} \qquad (1)$$

Equation (1) describes a simple VAM specification, including prior year test scores ($A_{i,t-1}$), student (X_{it}), and classroom (Z_{jt}) characteristics, and a teacher fixed effect (μ), where i indicates individual student, j indicates classroom, and t denotes time period or grade. There are many different ways to write VAM equations (see, e.g., Bitler et al., 2019; SAS Institute, n.d.) but all are generally examples of linear mixed models that decompose the variation in student test scores according to influences reflecting prior achievement, family, background and student characteristics, and teacher and school contributions. In addition to teacher effects on student learning gains, VAMs can also include terms to measure and assess the performance of schools (Goldstein & Foley, 2012; Goldstein & Leckie, 2008), students, and other units for which accountability testing may be a relevant policy objective (e.g., school districts, school types or sectors).

THE VAM THEORY OF CHANGE AND VAMS IMPLEMENTATION IN DIFFERENT JURISDICTIONS

The use of VAMs implies a specific theory of change in educational performance assessment. According to this theory, increased test-based accountability of teachers and schools leads to improved school leadership and greater teacher productivity. In turn, these elements drive increased student learning and hence improved student achievement. Aggregated up, improved student achievement lifts the system-level performance of the education system (Amrein-Beardsley, 2014: 11). By measuring teacher contributions to student learning and acting on those insights in relation to individual teachers, VAMs enable school administrators and educational decision makers to hold teachers and schools accountable and improve the overall performance of the education system by improving teaching quality.

The importance of teacher quality for student outcomes is uncontested, and teacher quality influences student learning more substantially than other in-school characteristics and processes (Clinton et al., 2016; Darling-Hammond et al., 2012). Improving the quality of teaching has also been shown to improve student outcomes within schools and at the level of whole education systems (Clinton et al., 2016). For these reasons, countries such as Singapore, the United Kingdom and Australia have all investigated using VAMs for evaluating education system performance. The OECD similarly carried out a research project across its member countries to show how to introduce VAMs-informed evaluation systems (OECD, 2008), and a recent review of empirical articles using VAMs to assess teacher and school effectiveness identified over 300 articles from 26 countries (Levy et al., 2019).

However, the United States stands out internationally for the comprehensiveness with which it adopted VAMs and for the direct way VAMs informed decision making in educational administration. The United States introduced VAMs following the Federal No Child Left Behind Act (2001) (Wong, 2012), which mandated standardised testing and statistical analysis of school and teacher performance across the

United States. Testing was limited to core subjects like English and maths for students in grades 3 to 8 and high school. In 2009, the Obama administration introduced Race to the Top (RTT) (Dragoset et al., 2016; RTT, 2009), which tied state education funding to the performance assessment of teachers and principals, using common assessment standards and better data systems.

Race to the Top allowed states to compete for federal grants, which were awarded according to criteria relating to teacher and principal quality, state education policies, standards and assessments, improving the performance of the lowest achieving schools and implementing data systems to support instruction (RTT, 2009). RTT implemented a competitive process to allocate federal funding specifically to allow the Obama administration to select proposals that relied heavily on data and evidence to demonstrate program effectiveness (Haskins & Margolis, 2014: 215). In response to RTT, many US states adopted highly aggressive policies of performance management, including the use of VAMs. The take-up of VAMs was also facilitated by readily available private sector proprietary products such as the SAS Institute's EVAAS (Education Value-Added Assessment) system (SAS Institute, n.d.) and widely available open source models. By 2014, approximately 40 US states were using VAMs for performance evaluation and monitoring, including for high-stakes decisions relating to teacher tenure, termination and pay. VAMs were only one component in teacher evaluation but were widely regarded as the most objective, especially in relation to other forms of evaluation such as supervisor assessments; objective evaluations were required under RTT legislation (Amrein-Beardsley, 2014; RTT, 2009).

Other countries have also adopted or considered the use of VAMs for school and teacher accountability testing and performance management. In 2008, the OECD published Measuring improvements in learning outcomes: Best practices to assess the value-added of schools (OECD, 2008). This report provided an overview of how to develop a system to undertake school-level Value-Added Modelling for performance assessment in OECD countries. Thirteen countries participated in the project, which documented policy and accountability objectives, technical and methodological descriptions of models, the use of incentives and sanctions for schools and teachers, the steps to implement a system of VAMs in member countries, and the already existing data and system resources in different countries that could go into the system. The OECD was motivated to undertake the project because VAMs were "state-of-the-art" methods for measuring school performance (OECD, 2008: 11).

The United Kingdom is one country to adopt VAMs for school-level accountability (OECD, 2008). The UK Government has published rankings of school performance (school league tables) annually since 1992 (Goldstein & Foley, 2012), to promote educational performance assessment, hold schools accountable and support school choice decisions by parents. School-level Value-Added Models were introduced in 2002 which did not adjust for student background characteristics, and in 2006, these models were replaced with "contextual Value-Added Models" which attempted to separate school contributions to student achievement from differences in student populations across schools associated with age, gender, socioeconomic status and ethnicity (Leckie & Goldstein, 2017). Five years later, the government discarded contextual Value-Added Models and replaced them with a simple Value-Added Model that did not adjust for student socioeconomic and demographic characteristics. The government explained this decision by arguing that adjusting for socioeconomic differences in VAMs also controlled for lower educational aspirations and achievement among disadvantaged students and thus risked entrenching educational disadvantage by allowing schools to neglect these factors (Leckie & Goldstein, 2017).

The 2011 decision also introduced another measure of student progress: Expected Progress or EP, a school-level indicator of the percentage of students who made "expected progress" in secondary school against standard levels of the national curriculum (Leckie & Goldstein, 2017). Finally, in 2016, the government scrapped the Expected Progress measure and introduced Progress 8 (P8), a value-added measure of student secondary attainment across eight subjects. The P8 VAM adjusts for some school intake differences in prior achievement, but not student socioeconomic or demographic characteristics (Leckie & Goldstein, 2019). Schools that fall below a floor standard on their P8 scores receive increased attention by the national Office for Standards in Education, Children's Services and Skills (Ofsted), and by local and regional authorities. In contrast, the highest performing schools are exempted from the following year's Ofsted inspections.

Different Australian states and territories have also considered VAMs for system performance management and accountability, most notably at the school level. Victoria has considered both student-level and school-level Value-Added Models (Department of Education and Early Childhood Development, 2011), with school-level models based on UK contextual Value-Added Models. The Victorian

Education Department views contextual VAMs as measuring school and system performance. From 2017, the Victorian Government has used the Differentiated School Performance Method (Department of Education and Training, 2019) to assess school performance. This measures the change in performance of individual schools over six domains: achievement, school climate, student attitudes, secondary performance, student engagement and participation, but is not based on VAMs (Department of Education and Training, 2019).

In 2014, the New South Wales Government's Centre for Education Statistics and Evaluation (CESE) developed and assessed a set of value-added measures for NSW Government schools (CESE, 2014). The models were developed to estimate school contributions to reading and numeracy from years 3 to 5, 5 to 7, 7 to 9 and 9 to 12, and used NAPLAN achievement data and NSW Higher School Certificate English and maths. The New South Wales models also adjusted for contextual student sociodemographic characteristics and were used to show the distribution of school contributions across the NSW Government school sector (CESE, 2014). The New South Wales analyses are specifically couched in terms of enabling schools to understand and improve their own performance, and the reports include a number of caveats about the interpretation and presentation of such models (Lu & Rickard, 2014). New South Wales does not use VAMs for performance assessment of individual teachers.

At the national level, the Australian Productivity Commission undertook an inquiry into the National Education Evidence Base in 2016 (Productivity Commission, 2016). The inquiry was designed to advise on a "national approach to collecting and using data for early childhood education and care and schools, and other information as relevant, to improve Australia's educational outcomes" (Productivity Commission, 2016: v). The inquiry endorsed the use of Value-Added Models for education outcomes and school achievement, and, like the CESE in NSW, argued that they could be used to identify high-performing schools, whose programs and practices might translate effectively to other schools (Productivity Commission, 2016). The inquiry also recommended that federal, state and territory governments should all use VAMs for education outcomes. Unlike the United States, however, the Productivity Commission did not recommend using them as teacher accountability measures.

In contrast to other countries, by 2014, about 40 US states were using VAMs as central elements in their education performance management systems to assess individual teacher accountability and compile the evidence for RTT funding (Close et al., 2018; Close et al., 2020). One of the more extreme examples of use occurred in the Houston Independent School district in Texas, which, in 2011, decided not to renew employment contracts for 221 teachers. In a number of cases, the school district attributed non-renewal to the teachers' poor performances on their individual VAMs results (Amrein-Beardsley, 2014: 25-49), assessed using the proprietary EVAAS system (SAS Institute, n.d.), which by 2014 was the most widely used VAM throughout the US (Amrein-Beardsley, 2014: 25).

The use of VAMs in the Houston Independent School District, and similar applications of the methods in other jurisdictions, eventually led to 15 lawsuits across the United States by teachers' unions and others. In most cases, VAMs were challenged on grounds including lack of reliability or consistency of VAMs measurement over time, lack of validity of the measures, bias due to results reflecting selective student populations (for instance, minority students, low-income students, English language learners, special education students), lack of fairness in that only maths and reading teaching were tested via VAMs, and lack of transparency. The lack of transparency in the use of VAMs for performance assessment was linked to the complexity of the models and results, and the sometimes confidential (because proprietary) nature of the VAMs algorithms and results. This meant that teachers could not understand the methods leading to their judgements or use results to improve teaching practice (Amrein-Beardsley, 2019; Richardson et al., 2019).

The Houston case, brought by the Houston Federation of Teachers, challenged VAMs for teacher accountability on constitutional grounds, namely, that the EVAAS algorithms the school district relied on for employment decisions were not meaningful to the teachers who were subject to those decisions. Teachers were thus denied due process (Amrein-Beardsley, 2019). The trial also highlighted that VAMs estimates of teacher effectiveness would be wrong if input data or underlying code were wrong, that correcting errors would potentially change results for all teachers being evaluated and that the school district had no controls or audit processes to check errors, or procedures or intentions to correct them if discovered (Amrein-Beardsley, 2019). In 2017, the case was settled with the Houston School District agreeing not to use VAMs to terminate teachers' employment if the VAMs scores could not be verified (Amrein-Beardsley, 2019).

Aside from litigation, using VAMs for teacher accountability generated much academic and policy controversy, which we consider next. But it also had unintended consequences that were not anticipated in the theory of change. These included school districts being unable to retain teachers, teachers refusing to teach high-needs students, refusing to collaborate with one another to share practice, and seeking teaching assignments where they believed they would score more highly (Johnson, 2015). The American Statistical Association (2014) and American Education Research Association (2017) released major statements calling into question the validity of the models and their use in high-stakes decisions, like teacher employment (American Education Research Association, 2017; American Statistical Association, 2014). Partly in response to these events, in 2016, the Obama administration signed in new federal legislation, the Every Student Succeeds Act (ESSA 2016), which allowed states and districts greater flexibility to design and implement teacher evaluation policies and systems (Close et al., 2018). By 2019, 15 US states were using VAMs state-wide for teacher evaluation (down from 40 in 2014), 21 were not, and the rest were using other methods which might include VAMs alongside other approaches.

THE ACADEMIC CONTROVERSY - WHAT HAS THE LITERATURE FOCUSED ON AND WHY DOES IT MISS THE POINT?

Using VAMs to assess teacher performance was not only contentious in educational policy making. For at least 15 years, academic research has also considered the use of VAMs, with the research addressing methodological (predominantly econometric and statistical) issues, policy issues, and critical analyses of the use of VAMs in education performance assessment (Evans & Lee, 2018). The methodological research has attempted to assess whether teacher contributions to student achievement exist and can be estimated statistically, and to assess issues such as statistical bias, reliability or stability of VAMs estimates over time, and validity (Andrabi et al., 2011; Bitler et al., 2019; Briggs & Domingue, 2011; Chetty et al., 2014a, 2014b; Durso, 2012; Evans & Lee, 2018; Reardon & Raudenbush, 2009; Stacy et al., 2018). The policy research has examined the use of VAMs in particular settings, particularly as these are linked to teacher quality and teacher performance assessment (Evans & Lee, 2018). Finally, the critical analysis examines the unintended consequences of using VAMs in real settings for teacher performance assessment (Evans & Lee, 2018 and for examples see Amrein-Beardsley, 2012, 2014; Amrein-Beardsley & Holloway, 2019; Amrein & Berliner, 2002, 2003; Cochran-Smith et al., 2016; Darling-Hammond et al., 2012; Pivovarova et al., 2016).

The methodological debate is well illustrated by Chetty et al. (2014a, 2014b) and Bitler et al. (2019). Chetty and colleagues specifically frame the controversy around the use of VAMs for evaluating teacher quality as a methodological issue—do VAMs capture real causal impacts of teacher quality on student learning or are they biased by non-random sorting of students to teachers (Chetty et al., 2014a)? Bias here is understood statistically to mean that the estimate of a teacher contribution differs systematically from the true value in a way that reflects how the model was specified and estimated. Using education and system data and methods to quantify bias, they show the bias is comparatively small. They then develop a method to assess bias in other settings using VAMs. A follow-up paper (Chetty et al., 2014b) further shows that teachers' contributions to student learning have long-term effects on students' outcomes as adults, such as earnings, university attendance and teenage birth. Having demonstrated the robustness of these research findings, they estimate the economic effects of policies which use VAMs to retain "high value-added teachers" and replace "low value-added teachers."

Bitler et al. (2019) specifically address the methodological problems of VAMs to question their use in high-stakes evaluation of teacher performance. By focusing on "an outcome that teachers cannot plausibly affect: student height" they use VAMs to estimate a teacher contribution and also identify whether it is due to students being sorted non-randomly by teachers or whether it is due to random sampling variation or noise (Bitler et al., 2019). They conclude it is due to random noise, and further test VAMs approaches that adjust for this, concluding that only one approach yields a correct adjustment. They conclude that because VAMs cannot plausibly distinguish real teacher contributions from random variation, policymakers and administrators should be cautious about using VAMs in practice.

These papers illustrate how the methodological debate around VAMs has been constructed. Proponents have tried to show that well-specified research demonstrates the high-level conclusion that teacher quality matters for student achievement, that VAMs can demonstrate this and therefore that VAMs should be used for teacher accountability and system improvement. Critics argue that methodological limitations in VAMs mean that inferences are biased, unreliable, or invalid. The methodological debate is particularly relevant to our argument because researchers who attempt to make a case "for" or "against" VAMs on methodological grounds typically do so using best-practice science, statistics, and econometrics. They proceed with high-quality research designs, quality-assured data, and robust and thoughtful modelling strategies, where they test and attempt to break conclusions. This approach to making a case "for" or "against" assumes a kind of methodological "kill-shot," whereby the best research decisively settles the argument about whether teachers contribute directly to student learning, and the decision to use VAMs for teacher accountability and system performance follows logically and straightforwardly.

Advocates and critics of the use of VAMs for high-stakes decision making both assume that a central issue is whether these models can reliably estimate teacher effects on student learning. However, even if research showed VAMs unambiguously measure teacher effectiveness, and teacher effectiveness was a primary driver of student learning, simply rolling out VAMs in education performance monitoring and high-stakes decisions about systems, schools, teachers and students would not be justified. As soon as policymakers and education administrators propose using VAMs to make operational decisions about individual teachers, students, or schools (i.e., administrative business decisions such as whether to hire, confirm or dismiss an individual teacher, or progress a student), other research and ethical considerations come into play that mean that the weight of scientific evidence under controlled conditions is not sufficient to justify using VAMs in operational deployment for education performance assessment.

Critical policy research (e.g., Amrein-Beardsley & Holloway, 2019) that argues against VAMs on the basis of research evidence also fails to address some pertinent research and ethical issues. Amrein-Beardsley and Holloway (2019) argue that the belief that teacher quality is the most important in-school factor leading to student achievement rests on assumptions including the following:

1. Teachers are the most important in-school factor in student achievement.
2. Good teaching comes from enduring qualities that teachers carry with them from one year to the next, regardless of context.
3. US public school teachers are ineffective and hinder progress.
4. VAMs will improve what is wrong with faulty teacher evaluation systems.
5. Performance assessment can be used to reform schools.
6. VAMs can be used and understood by teachers to improve instruction.
7. VAMs provide methodologically robust measures of the educational value added by teachers.

By examining the evidence, Amrein-Beardsley & Holloway (2019) attempt to show that VAMs rest on flawed theory, methods and substantive framework, and so these assumptions fail. We agree with much of their argument, but the implication of this approach is that if the above assumptions stand, using VAMs for operational decision-making is legitimate. We reject this implication. Even if the assumptions could cogently be substantiated, additional questions remain in using VAMs in high-stakes operational decisions in real-world settings such as school districts or education systems.

The American Statistical Association made this point in its 2014 Statement on the use of VAMs. While the Association endorsed the "wise use" of data, statistical models and experiments to improve educational quality, it also distinguished low-stakes applications, such as evaluating policy and program effectiveness, from high-stakes applications, such as deciding teacher, pay, tenure or dismissal. In low-stakes applications, VAMs could help improve educational quality by allowing teachers and schools to assess students' relative performance against other similar students and adjust teaching practice and professional development accordingly. But even in low-stakes cases, using VAMs meaningfully requires careful attention to issues such as data quality, measurement reliability and validity, which variables are included in models, potential bias associated with different student sub-populations, and taking account of model precision and uncertainty (American Statistical Association, 2014).

In high-stakes cases, the Association crucially recognised that the impact of VAMs on education quality depended not only on the statistical and methodological characteristics of VAMs but also on the way models, data, systems and accountability frameworks were operationally deployed in real settings (American Statistical Association, 2014). The academic debate on VAMs misses the point because it does not recognise the crucial difference between high and low-stakes applications of VAMs, namely, that in high-stakes cases, the impact of VAMs on education quality depends not only on the veracity of data and models, but also on how results are derived and used to inform high-stakes decisions.

In high-stakes applications, there is a stronger ethical case to address, a different burden of proof on evidence, and a set of implementation issues to consider that allow transparency, fairness, and accountability—in other words, due process—to those subject to decisions. The lack of due process in operational deployment is precisely the basis on which the use of VAMs has been overturned in the courts in the United States.

Several related fields provide indications of how researchers might approach these issues. Clinical medicine, public health, and health research more generally recognise that the research to develop clinical solutions is not sufficient to implement those solutions in real-world settings. In these fields, implementation research or implementation science aims to clarify how to implement (deploy) evidence-based health solutions in new contexts and populations (Gopichandran et al., 2016). For example, the clinical approach to developing a new vaccine involves basic science, clinical research, vaccine development, and testing for safety and efficacy. Research takes place under strictly controlled conditions and includes designs such as laboratory science, randomised clinical trials, and matched case control studies, and ethical principles relate to harm minimisation and beneficence to study participants. Broader issues such as how to integrate the new vaccine into existing healthcare systems with different sub-populations and contexts are left unaddressed (Gopichandran et al., 2016).

Deploying a new vaccine inside a health system raises additional issues that are not examined by clinical researchers undertaking vaccine development. These include appropriateness and acceptability for different sub-populations, feasibility of implementation across settings and sub-populations, equity, cost-effectiveness and so on. Implementation research addresses these issues. Implementation research can help inform effective deployment by providing evidence relating to health promotion strategies, the conditions and limits of social acceptability of the vaccine, ways to address vaccine hesitancy, integration into the broader health system, logistics for mass vaccination, where bottlenecks or gaps in delivery are, and what population or environmental conditions are necessary for successful deployment (Gopichandran et al., 2016).

These kinds of questions also require different disciplinary approaches, research designs and ethical perspectives from clinical research. Whereas research designs for clinical development emphasise controlled (often experimental) procedures in controlled environments (e.g., laboratories, clinical trials) and yield "gold-standard" research results based on purpose-built research artefacts, implementation research uses a wider range of research designs (including non-experimental, qualitative, mixed and participatory methods), addresses other questions and takes place in real settings. While clinical research focuses on harm and benefit to direct participants, implementation research considers holistic impacts on overall systems, including potentially unintended consequences for others who may or may not be the intended beneficiaries of the implementation (Gopichandran et al., 2016).

The distinction between clinical research and implementation research is informative for the deployment of VAMs in school systems in the United States. The academic debate around VAMs proceeded as if the key issue was whether VAMs could reliably estimate the contribution individual teachers made to student learning, under the tightly controlled but purpose-specific conditions of academic research. If VAMs were academically defensible in this setting, that would justify building test-based accountability systems using VAMs and deploying them in actually existing school systems under operational conditions. The high-stakes implementation or deployment issues noted by the American Statistical Association were not recognised as issues requiring additional research, or ethical commitments to transparency, accountability, system-level consequences, and system-level harm minimisation.

An area somewhat closer to this case study that recognises research and ethical issues around real-world deployment is algorithmic decision making in public policy or commercial activity using approaches such as machine learning and artificial intelligence more broadly. Artificial intelligence and machine learning methods have catalysed very substantial research literatures, frameworks and general principles to understand the “ethical and societal implications” of using data and algorithms to inform decision-making in business, policy and practice. Ethical and societal implications relate to understanding how using these approaches might impact parts of society and enhance or threaten societal values (Whittlestone et al., 2019). This research encompasses the social sciences, law, philosophy, and computer science, among other disciplines (Jobin et al., 2019; Lo Piano, 2020). While the literatures have not settled on a single set of ethical concerns or principles relating to the use of data and algorithms, common themes include fairness, accuracy, accountability and transparency (Lo Piano, 2020) and transparency, justice and fairness, non-maleficence, responsibility and privacy (Jobin et al., 2019). The key ethical principles seem to be approaching those found in health and medical research (Mittelstadt, 2019) and align with issues we have noted in implementation research. The attention to ethical issues is also driving research into the technological capabilities and limitations of algorithmic decision making using machine learning, the uses and impacts of these technologies, including their application in real-world settings, and the perspectives of different communities, stakeholders and groups who are affected by these methods. Where the research around VAMs failed largely to consider the ethical and research questions that were unresolved in using VAMs for administrative and operational decision making, the emerging research on algorithmic decision making using machine learning and AI increasingly centralises these concerns (Lo Piano, 2020).

WHAT DO WE NEED TO GO FORWARD?

The debate around the use of VAMs highlights many issues that are relevant to the use of big data for social policy. First, it is critical to understand and specify “use cases” or applications when identifying how research can support policy. How and for what purposes are big data intended to be used? Building a high-level systemic policy around a research evidence base is a different application than using the results of research or algorithmic decision making to make day-to-day operational decisions that directly affect individuals, whether these are employment related, or tied to other issues such as whether they receive a service or payment or have breached or infringed rules. Both these examples differ from a more abstract (academic) question, such as: Can VAMs reliably estimate individual teacher contributions to student learning and allow teachers to be ranked on these contributions? In all three types of applications, the scientific and research evidence needs to be credible and robust, but evidence that VAMs are scientifically valid for drawing academic research conclusions does not necessarily imply they should be used for educational policy. The research evidence base that supports policy is also not sufficient to tell us if or how VAMs should be deployed operationally for high-stakes decision making.

Using data and algorithms directly to decide on the fates of individuals raises implementation challenges and associated research questions and ethical issues. The challenges can be addressed through research, but only if researchers understand some of the complexity of real-world implementation and focus on appropriate research questions with appropriate research designs. For high-stakes operational deployment, researchers need to attend to system-level interactions and intended and unintended consequences within the broader system. Implementation research also requires an ethical lens that goes beyond principles of harm minimisation and beneficence to study participants, narrowly conceived.

Operational deployment or implementation requires more than scientifically credible findings about the effectiveness or efficacy of a proposed approach under the tightly controlled conditions of academic research. This is a necessary but insufficient condition for implementation. When operational data, systems, and algorithms are not accurate, decisions are not transparent and contestable, governance arrangements do not allow for errors to be corrected, or effective accountability and redress, research, policy and procedural corrections that address these issues are also required. Without an evidence-informed approach, implementation can lead to unintended and unforeseen consequences, wasted resources and social harm.

As social science research is increasingly directed to informing the development of solutions to real-world challenges, researchers who want to contribute to those solutions need to better understand how and under what conditions their research will be used. Much of the research around the use of VAMs for educational performance assessment failed to recognise or acknowledge this. The failure to recognise the need for implementation research and appropriate ethical deployment in this case study says much about boundaries between disciplinary knowledge and between researchers, policymakers and other users of research.

REFERENCES

ABS (2012). Wage price index: Concepts, sources and methods. https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/6351.0.55.001

American Education Research Association (2017). Knowledge to action: Achieving the promise of equal educational opportunity. 2017 AERA Annual Meeting, San Antonio, TX.

American Statistical Association (2014). ASA statement on using value-added models for educational assessment.

Amrein, A. L., & Berliner, D. C. (2002). An analysis of some unintended and negative consequences of high-stakes testing. Education Policy Research Unit, Education Policy Studies Laboratory, College of Education.

Amrein, A. L., & Berliner, D. C. (2003). The testing divide: New research on the intended and unintended impact of high-stakes testing. Peer Review, 5(2), 31-32.

Amrein-Beardsley, A. (2012). Recruiting expert teachers into high-needs schools: Leadership, money, and colleagues. Education Policy Analysis Archives, 20(27), 1-23.

Amrein-Beardsley, A. (2014). Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability. Routledge.

Amrein-Beardsley, A. (2019). The Education Value-Added Assessment System (EVAAS) on trial: A precedent-setting lawsuit with implications for policy and practice. eJEP: eJournal of Education Policy.

Amrein-Beardsley, A., & Holloway, J. (2019). Value-added models for teacher evaluation and accountability: Commonsense assumptions. Educational Policy, 33(3), 516-542.

Andrabi, T., Das, J., Khwaja, A. I., & Zajonc, T. (2011). Do value-added estimates add value? Accounting for learning dynamics. American Economic Journal: Applied Economics, 3, 29-54.

Australian Council for Educational Research (2020). Trends in international mathematics and science study. https://doi.org/https://www.acer.org/au/timss

Australian National Audit Office (ANAO) (2012). National partnership agreement on literacy and numeracy. https://www.anao.gov.au/work/performance-audit/national-partnership-agreement-literacy-and-numeracy

Bitler, M., Corcoran, S., Domina, T., & Penner, E. (2019). Teacher effects on student achievement and height: A cautionary tale. National Bureau of Economic Research Working Paper Series, No. 26480. https://doi.org/10.3386/w26480

Briggs, D., & Domingue, B. (2011). Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles unified school district teachers by the Los Angeles Times. Boulder, CO: National Education Policy Center.

Brown, M., McNamara, G., & O'Hara, J. (2016). Quality and the rise of value-added in education: The case of Ireland. Policy Futures in Education, 14(6), 810-829.

CESE (2014). Centre for Education Statistics and Evaluation (CESE), New South Wales Government, Australia. https://www.cese.nsw.gov.au/

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. American Economic Review, 104(9), 2593-2632. https://doi.org/10.1257/aer.104.9.2593

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. American Economic Review, 104(9), 2633-2679. https://doi.org/10.1257/aer.104.9.2633

Clinton, J., Anderson, M., Dawson, G., Dawsaon, A., Bolton, S., & Mason, R. (2016). Teacher effectiveness systems, frameworks and measures: A review. Melbourne Graduate School of Education.

Close, K., Amrein-Beardsley, A., & Collins, C. (2018). State-level assessments and teacher evaluation systems after the passage of the Every Student Succeeds Act: Some steps in the right direction. http://nepc.colorado.edu/publication/state-assessment

Close, K., Amrein-Beardsley, A., & Collins, C. (2020). Putting teacher evaluation systems on the map: An overview of states' teacher evaluation systems post–Every Student Succeeds Act. Education Policy Analysis Archives, 28, 58.

Cochran-Smith, M., Stern, R., Sánchez, J. G., Miller, A., Keefe, E. S., Fernández, M. B., Chang, W. C., Carney, M. C., Burton, S., & Baker, M. (2016). Holding teacher preparation accountable: A review of claims and evidence. National Education Policy Center.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. Phi Delta Kappan, 93(6), 8-15.

Department of Education and Early Childhood Development (2011). Annual report 2010-11. https://www.education.vic.gov.au/Documents/about/department/201011deecdannualreport.pdf

Department of Education and Training (2019). Victoria State Government, Australia. https://www.education.vic.gov.au/Pages/default.aspx?

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Tanenbaum, C., Giffin, J., & Upton, R. (2016). Race to the Top: Implementation and relationship to student outcomes: Executive summary (NCEE 2017-4000). https://ies.ed.gov/ncee/pubs/20174001/pdf/20174000.pdf

Durso, C. S. (2012). An analysis of the use and validity of test-based teacher evaluations reported by the Los Angeles Times: 2011. National Education Policy Center.

Evans, C. M., & Lee, J. C. (2018). Value-added assessment of teacher preparation programs in the United States: A critical evaluation. Assessment in Education: Principles, Policy & Practice, 25(5), 519-539. https://doi.org/10.1080/0969594x.2016.1255180

Everson, K. C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? Review of Educational Research, 87(1), 35-70.

Everson, K. C., Feinauer, E., & Sudweeks, R. (2013). Rethinking teacher evaluation: A conversation about statistical inferences and value-added models. Harvard Educational Review, 83(2), 349-370.

Goldstein, H., & Foley, B. (2012). Measuring success: League tables in the public sector. British Academy Policy Centre report.

Goldstein, H., & Leckie, G. (2008). School league tables: What can they really tell us? Significance, 5(2), 67-69.

Gopichandran, V., Luyckx, V. A., Biller-Andorno, N., Fairchild, A., Singh, J., Tran, N., Saxena, A., Launois, P., Reis, A., & Maher, D. (2016). Developing the ethics of implementation research in health. Implementation Science, 11(1), 161.

Haskins, R., & Margolis, G. (2014). Show me the evidence: Obama's fight for rigor and results in social policy. Brookings Institution Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school? Educational Researcher, 44(2), 117-126. https://doi.org/10.3102/0013189x15573351

Kleinhenz, E., & Ingvarson, L. (2004). Teacher accountability in Australia: Current policies and practices and their relation to the improvement of teaching and learning. Research Papers in Education, 19(1), 31-49.

Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992-2016: ‘Contextual val-added,’ ‘expected progress’ and ‘progress 8.’ British Educational Research Journal, 43(2), 193-212.

Leckie, G., & Goldstein, H. (2019). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. British Educational Research Journal, 45(3), 518-537.

Levy, J., Brunner, M., Keller, U., & Fischbach, A. (2019). Methodological issues in value-added Smodeling: An international review from 26 countries. Educational Assessment, Evaluation and Accountability, 31(3), 257-287.

Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward. Humanities and Social Sciences Communications, 7(1), 1-7.

Lu, L., & Rickard, K. (2014). Value added models for NSW Government schools. https://www.cese.nsw.gov.au/images/stories/PDF/VAPaper_v3-1Final.pdf

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507.

OECD (2008). Measuring improvements in learning outcomes: Best practices to assess the value-added of schools (9264050256). https://www.oecd.org/pisa/

Pivovarova, M., Amrein-Beardsley, A., & Broatch, J. (2016). Value-added models (VAMs): Caveat emptor. Statistics and Public Policy, 3(1), 1-9. https://doi.org/10.1080/2330443x.2016.1164641

Productivity Commission (2016). Report on government services 2016.

Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. Education Finance and Policy, 4(4), 492-519. www.jstor.org/stable/educfinapoli.4.4.492

Richardson, R., Schultz, J. M., & Southerland, V. M. (2019). Litigating algorithms 2019 US Report: New challenges to government use of algorithmic decision systems. AI Now Institute. https://ainowinstitute.org/litigatingalgorithms-2019-us.html

Rowe, E. (2017). A penalizing system of accountability in Australia exacerbates equity gaps in education. World Education Blog, 2017/18 GEM Report. https://gemreportunesco.wordpress.com/2017/12/05/a-penalizing-system-of-accountability-in-australia-exacerbates-equity-gaps-in-education/

RTT (2009). Race to the Top (RTT). https://www2.ed.gov/programs/racetothetop/factsheet.html

SAS Institute (n.d.). SAS® EVAAS® for K-12 statistical models. SAS EVAAS White Paper. https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf

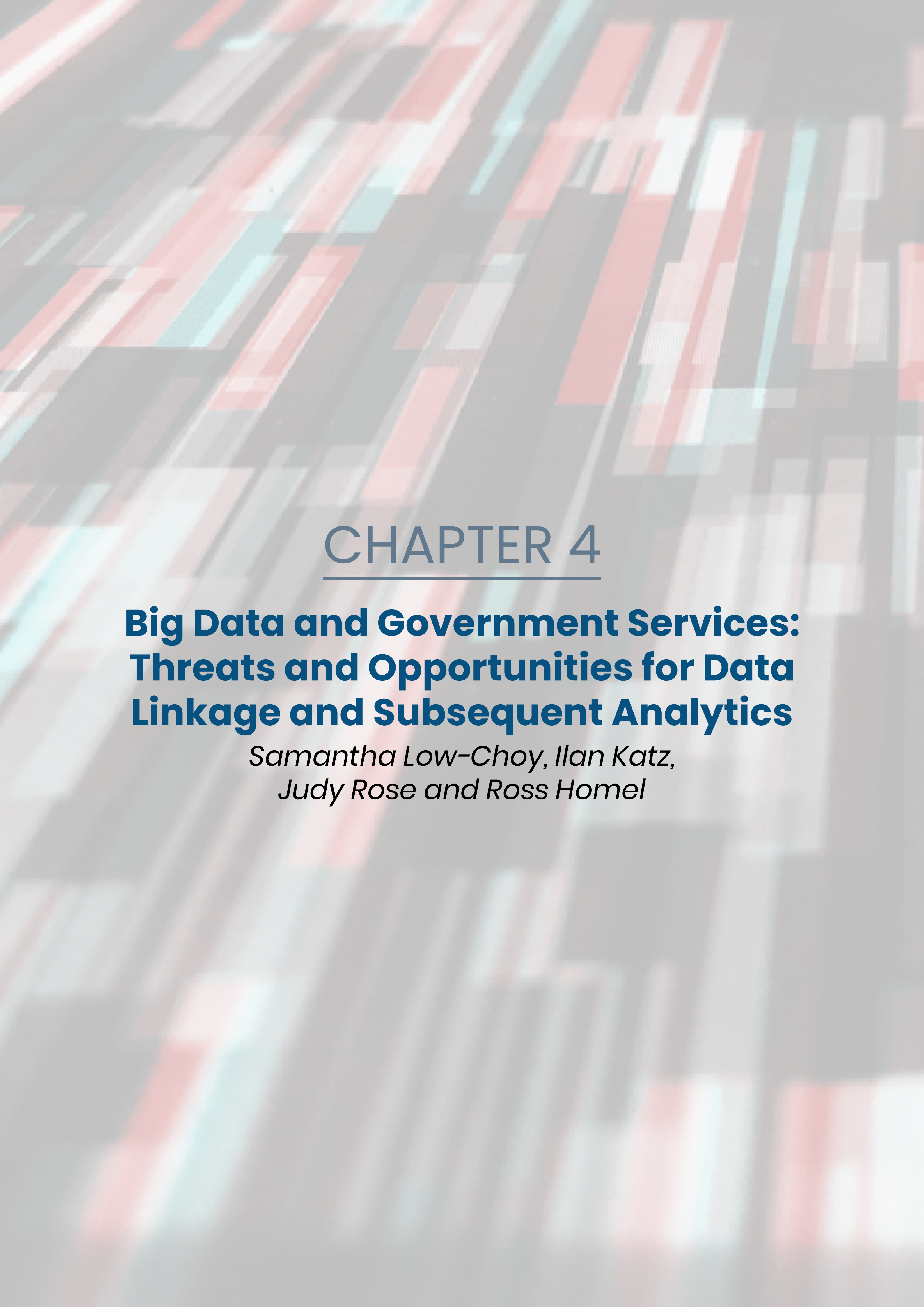
Stacy, B., Guarino, C., & Wooldridge, J. (2018). Does the precision and stability of value-added estimates of teacher performance depend on the types of students they serve? Economics of Education Review, 64, 50-74.

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research. https://www.nuffieldfoundation.org/wp-content/uploads/2019/02/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf

Wong, K. K. (2012). The implementation of the No Child Left Behind Act: Toward performance-based federalism in US education policy (Chapter 3). Australian Productivity Commission and Forum of Federations: Benchmarking in federal systems.

The World Bank (2018). Education data & measurement. https://www.worldbank.org/en/topic/education/brief/education-data-measurement/

Wright, D. B. (2017). Using graphical models to examine value-added models. Statistics and Public Policy, 4(1), 1-7. https://doi.org/10.1080/2330443x.2017.1294037



CHAPTER 4

Big Data and Government Services: Threats and Opportunities for Data Linkage and Subsequent Analytics

Samantha Low-Choy, Ilan Katz,
Judy Rose and Ross Homel

Big Data and Government Services: Threats and Opportunities for Data Linkage and Subsequent Analytics

INTRODUCTION

This chapter starts a series of three (with Chapters 5 and 6) that focus on data linkage applied to big data within the context of social policy. Data linkage brings together information from different sources about the same person to create a new, richer dataset that can be used for research or policy (United Nations Statistical Office, 1991). The complex process of data linkage (OAIC, 2014) involves the technical challenges of data matching and subsequent data analysis (Tahamont et al., 2020). Within the linkage process, data matching refers to the technical effort that ensures that individual identities are faithfully matched when combining datasets (Harron et al., 2017). Data analytics enables big data to be exploited, reshaping raw data into information crucial to understanding and decision-making (Mikalef et al., 2020). Given these powerful abilities, data linkage is now widely used by government agencies throughout Australia, including those with a key role in social policy research and development. Specialist institutions conduct data linkage and manage the complex process. For instance, the Australian Institute for Health and Welfare and the Australian Institute for Family Studies note as part of their mandate: “Our linking of data sets provides valuable new information for research and policy-making, in a secure, privacy-preserving manner” (http://www.aifs.gov.au/our-workresources/data-linkage). In these three chapters, the analysis and discussion draw out some of the ways in which data linkage can support richer use of information, with examples from recent policy initiatives, but importantly, they also identify the risks that are involved.

Emergence of big data

Big data has been portrayed as “the next frontier for innovation, competition, and productivity” (Mikalef et al., 2020: 1). Definitions of “big data” often refer to popular notions of the 3 Vs, of large volume, wide variety, with change at great velocity (e.g., De Mauro et al., 2016). While there are different types of big data, the focus in our three chapters (4-6) is on administrative data, which is highly relevant to the social sciences and to social policy. We consider administrative data as “derived from the operation of administrative systems (e.g., data collected by government agencies for the purposes of registration, transaction and record keeping)” (Elias 2014, as cited in Connelly et al., 2016: 2). Administrative data contains information that identifies individuals, as well as other highly confidential information (e.g., income or health) and therefore presents specific challenges (Harron et al., 2017). However, linked administrative data provide a useful resource for examining social policy issues (Connelly et al., 2016: 5), and forms an essential component of any analysis of the impact of big data more generally.

It is now axiomatic that data is the “oil of the 21st century” (credited to the mathematician Clive Humby). Like oil in the 19th and early 20th centuries, data has now become recognised as a valuable commodity which has to be mined and refined in order to be useful. In particular, big data is now often seen as “a powerful new resource for social science research” (Connelly et al., 2016: 1), offering many benefits to the public, since “there is a strong argument also for evidence-based government decisions and that evidence can be found through analysing huge amounts of data” (Galloway, 2017: 91). Like oil, data also has a dark side, benefiting some groups while disadvantaging others. Data has to be “extracted” from individual people, who may or may not benefit from the use of their data.

Risks and benefits of big data, linkage and analytics

This dilemma of the potential benefits and risks of big data has been significantly highlighted by recent debates in Australia around the response to the COVID-19 pandemic. Smartphones can be used to track people’s movements, to facilitate contact tracing to identify the source of each infection. Before participating in smartphone surveillance, each person therefore needs to balance health benefits (for themselves and the public) against consequences such as “surveillance creep,” which may impinge on their privacy (Garrett et al., 2021). The resulting surveillance data is crucial to informing responses to the epidemic, which relies on tracing infection sources, understanding delays in symptoms and/or reporting, and predicting the path of the epidemic (Price et al., 2020: 8). Beyond COVID-19, big data provide implications for future expansion of a national “One Health” information system (Johnston with PHAA, 2020: 442). The debates regarding COVID-19 “apps” in various countries, including Australia, illustrate the enormous potential benefits of data collection, linkage and subsequent analysis. However, they also highlight the legitimate concerns of the public about the use of big data, especially data collected by governments (Elkhodr et al., 2021). More generally, it has become well accepted that data linkage is not solely a technological concern, and has often become intertwined with infrastructure, social and ethical challenges (Harron et al., 2017).

The issues concerning data linkage necessarily address the analytics required to reshape the data into useful information. As noted by McCarthy and colleagues (2019: 15): "Data itself does not have value. Rather, it is what you do with the data that adds value." Indeed, all data must be analysed and interpreted by human beings, so it is misguided to assume that the data themselves will provide explanations, either for any individuals or a groups of individuals. Moreover, the size of big data makes it unwieldy to work with. Thus data analytics (including predictive analytics) have evolved to make use of algorithms that reveal patterns and trends in big data. However, care is needed when using data to find new patterns rather than confirming suspected patterns (Debackere, 2016), to avoid spurious, nonsensical or even dangerous conclusions. Connecting the data to existing knowledge bases and theories helps ensure that analytics are explainable, an important concern of the past decade which relates to people's rights (Kaminski, 2019), and "invokes a host of medical, legal, ethical, and societal questions that require thorough exploration" (Amann et al., 2020).

Data linkage allows research to be conducted on vulnerable populations to identify and track individuals who fall between the cracks of the human service systems and provide evidence to support them more effectively (e.g., homelessness during COVID-19, Herde et al., 2020). Linked data can also be used in more punitive ways, including to identify miscreants and rule breakers and to sanction them. One high-profile example of that usage has been the Robodebt initiative in Australia. Robodebt provides a stark reminder of how big data (through linkage and analytics) can be used for punitive purposes, for compliance, whilst aiming for beneficial purposes and for ensuring integrity of the social welfare system. The way this controversy unfolded highlights a number of issues and concerns about how data is used (and abused) by government agencies. Hence, this chapter adopts Robodebt as a central example that both introduces and synthesises the issues involved in data linkage and the analytics of big data more generally. Of interest is how experts interviewed, coincidentally at the time that Robodebt was unfolding, considered the practical benefits and risks of data linkage when analysing big data for the delivery of human services more generally.

Context of this study

All three chapters (4, 5, and 6) draw on the same set of interviews with experts, yet analyse different subsets of these interviews. These analyses adopt distinct methodologies, in order to examine different issues regarding data linkage. This chapter considers the experts' views on data linkage in practice, examining these within the context of the Robodebt affair, which escalated around the time of interviews. The subsequent analyses in Chapter 5 present a general thematic analysis of the issues raised by experts, culminating in a detailed analysis (with mixed methods) of the theme of privacy-trust-sharing, which was the theme most frequently mentioned by experts. Chapter 6 considers data linkage as a means of evaluating place-based initiatives for children and families.

This chapter examines some of the key challenges and opportunities for using big data (linkage and analytics), at a time when the uses and abuses of data are rapidly emerging as public policy issues in Australia. We start with a note on the research method underlying the interviews with experts on data linkage. Perceptions gained from interviews are organised into two sections: on general perceptions on big data and data linkage, and misuse of linked data. This sets the scene for a section on Robodebt: Recounting the Issues and Controversy. These are triangulated with the literature and with evidence of how experts were assimilating lessons from Robodebt (3 years before its resolution).

A NOTE ON RESEARCH METHOD

Strategy

Research on linking big data has tended to use case studies to facilitate nuanced analyses in specific contexts (Bloothooft & Schraagen, 2015), or has aimed for a broader snapshot using surveys (McNamara et al., 2018). The analyses for Chapters 4, 5 and 6 have a somewhat different aim and therefore take a different tack, by interviewing experts who have practical experience with data linkage in policy and research. Beyond these studies on data linkage, expert knowledge has made valuable and unique contributions in many fields of enquiry, including at the complex interface of social ethics and technology, for example, management of public infrastructure (Garthwaite & O'Hagan, 2000) and designing clinical trials (Best et al., 2020). Integrating data, theory, and other sources of knowledge, experts have refined abilities to distil those elements that differentiate important aspects, informed by their experience in many situations (Caley et al., 2014). Moreover, conscious of their reputations in their fields, experts can provide judgments that carry considerable weight. The experts' views are triangulated with the literature to provide insight into practices and attitudes to data linkage, at a time when these were evolving as the Robodebt affair unfolded.

Experts

Underpinning the three chapters lies a "core" set of 10 interviews, with high-level experts on data linkage: nine holding positions within government and one a career university researcher. For this chapter, four additional interviews were added: two with members of the research team, who had extensive experience in policy related research; and two extra interviews with high-level experts from government, with specific knowledge of the Robodebt affair. The latter two interviews focused specifically on historical and recent data linkage practices in the administration, monitoring and reform of the social security system. Their contributions are discussed solely in this chapter (i.e., not in Chapters 5 and 6), within the context of the Robodebt experience. Contributions from the two research team members are used in this chapter (and also in Chapter 6), since they add useful input on how linked data has been used to evaluate place-based initiatives for children and families. Altogether, analyses for this chapter draw on a total of 14 interviews.

In terms of experience, 8 of the 14 experts interviewed had worked in government for most or some of their career, three had careers focused in the university sector, and three had worked in both government-based policy contexts and academia, at different times. Most participants had significant involvement with data—in the form of linkage of "big" administrative data—in their day-to-day work roles. Only one expert appeared to have little hands-on engagement with data in their current role, although their perspective was valuable, as their role was mainly concerned with how personal data was used.

Interviews

The interviews were conducted during 2018 and 2019 and followed a semi-structured interview protocol (as described by, e.g., Leech, 2002). As such, the questions were not followed rigidly, so that the interviewers were able to shape the conversation to elicit the information of interest, whilst also pursuing topics raised by the experts. The interview questions explored the following issues:

- How the experts defined terms, including big data and data linkage;
- The current state of big data/data linkage in Australia;
- Examples of data linkage in social services and which data could be usefully linked;
- The benefits of data linkage in the context of big data (especially administrative data);
- The challenges of data linkage in the context of big data (open-ended); and
- Innovative uses of big data and data linkage in national and international contexts.

A full set of questions is detailed in Rose et al. (forthcoming). Interviews were conducted around the time that the Robodebt crisis was still undergoing debate and resolution. Interviews followed a semi-structured protocol (described in Chapter 5) and were conducted either face-to-face, by telephone or video conferencing. All were recorded, with the participants' consent, and transcribed verbatim.

Analysis

For this chapter, the analytical approach adopted a narrative discourse analysis (Stickley et al., 2007) of the topics raised during interview on the experiences, issues and perceptions of big data in terms of linkage and analytics. A distinguishing feature of narrative discourse analysis is that it views language as constitutive of experiences of working with data linkage in the context of big data, rather than attempting to identify themes that represent all views across all experts. This holistic approach contrasts with the more granular approach of subsequent analyses in Chapter 5, which dissect each expert's interview into blocks of text, each coded as addressing various themes. The main aim of the narrative analysis is to describe these experts' experiences with data linkage in practice, as a basis for comparison with the lessons learned from challenging data linkages, exemplified by the Robodebt affair.

When viewed in the context of the three related analyses (Chapters 4-6), this first analysis (Chapter 4) provides an overview of issues of data linkage in practice, in comparison with the sharper focus on a few issues (privacy-trust-sharing) of the second analysis (Chapter 5). This overview also provides a baseline for the narrower focus of the third analysis reported in Chapter 6, on a particular use of data linkage, to support evaluation of place-based initiatives to enhance wellbeing of children and families.

PERCEPTIONS OF "BIG DATA" AND DATA LINKAGE

Interviews started by seeking general perceptions on Big data, linkage and analytics before addressing the Broad benefits of data linkage, as a foil to their final comments regarding the Current state of data linkage in Australia. At this point, we consider two key issues that determine whether data linkage is possible: Technical ability and technological infrastructure to undertake data linkage, and the human infrastructure component of Technical skills and capacity.

Big data, linkage and analytics

To establish a baseline for the interview, experts were asked to outline what they understood by the terms "big data" and "data linkage." They tended to characterise big data as aggregate, society-level data, with some referring to "the 3Vs" of velocity, volume and variety (mentioned in the introduction to this chapter). Examples of big data cited by those interviewed included: the household census; retail pricing data; as well as internet-derived data, obtained from sources such as Google searches, social media, or an online national database called "My Health Record." Big data was generally characterised as dynamic, "coming in thick and fast." One expert (#3) felt that the term "big data" was about more than just the data, also encompassing the "data sources, the analytical techniques and the technology" needed to bring it all together.

There was an agreement amongst the experts that the main value of big data is enhanced by linking different datasets. Data linkage was characterised as pertaining to an individual ("whether that's a person or an organisation," (Expert #1) and bringing together all the records that belong to that individual "to create new rich and useful datasets" (Expert #1). Experts noted that data linkage often focused on more static, administrative data, and one questioned whether the linking of several datasets could create big data. This accords with technical definitions (as noted in the introduction to this chapter).

One expert spoke of an evolution of sorts in how people talk about data, making two main points, first, that the term "big data" is somewhat dated, and that over the last few years, the focus has shifted to "smart data", which refers to more than the volume of data, rather, what is done with it. Second, the focus has shifted from big data to data linkage: knowing what to link can lead to real breakthroughs. These are two of the few examples referring to data analytics (as prompted by the first interview question); most experts focused on recounting their understanding and experiences with big data and data linkage. Overall, there was a sense, amongst the experts consulted, that although the environment with regard to data linkage was improving, there was still "a way to go" (Expert #6).

Broad benefits of data linkage

Several experts gave examples of the benefits of data linkage in a range of contexts. Reflecting the fact that the majority of experts had held policy roles in government, the examples cited included using linked data to: provide "proactive" (Expert #12) and individualised responses in terms of healthcare, generate an evidence base for better public policy and to inform better program design, and provide a "rapid response" (Expert #2) to policy questions.

Regarding effective individualised responses, one expert described how the agency gave de-identified data back to general medical practitioners (GPs), who then re-identified the data, with permission from individual patients, to help manage patient health. It was emphasised that "at no point do we see identified data, but the packages that we give to the GP, they are able to extract [based on] the guidance that we give them; they're able to extract the patient" (Expert #12). This expert reported that any costs incurred in the data analysis stage were outweighed significantly by subsequent health cost savings. They also distinguished the use of data linkage for service provision, where it is possible to identify individuals in real time (with the individual's consent) and intervene appropriately to support their health. This contrasts with research and policy development context where individual identification is not ethical, where data that is retrospectively obtained (without explicit consent for specific linkage) must be de-identified.

At the broader scale of policy, one expert referred to the importance of data linkage in creating a better evidence base for program design and service interventions, and to prevent spending on programs that don't work. This was echoed by another expert, who spoke of the importance of using data linkage to scope the size of a population that might be eligible for an intervention, such as an early intervention program to prevent child maltreatment in particular sub-populations. Data linkage could help profile that sub-population—location, size and so on—to assess who would be eligible for the intervention. This active, targeted approach to enlistment contrasted with the more usual passive, untargeted approach of enrolling people who walk through the door of a service; only a small proportion would be eligible, and thus the intervention wouldn't "shift the population curve of the problem that you're trying to reduce" (Expert #2). Data linkage was therefore considered an important component of a more strategic approach to addressing the needs of vulnerable populations.

Regarding rapid response for policy, a university-based expert working with "novel linked data platforms" described how their team can now respond to data queries from government within a few days or a week "because that's the policy cycle" (Expert #2). This expert commented that this provides government access to analytic skills and data not otherwise available.

> *What we've started to realise is that the skillsets that sit within universities, especially around analysing data correctly … often those higher-level analytic skills are not available in government. If they are available, they sort of only sit in Treasury or something like that. That's really not something we ever planned, but that role has emerged. (Expert #2)*

Noting that, for researchers and universities, this capacity for rapid response is not the usual pace, this expert reported this kind of service is a "legitimate role."

Current state of data linkage in Australia

Towards the end of the interviews, experts were asked broadly for their views on the "current" state of big data in Australia. However, because these interviews were conducted between October 2018 and August 2019, they did not reflect any changes that have occurred since that time, such as the resolution of Robodebt. Even so, several key points were made by experts.

The absence of a national approach to data linkage was echoed by one expert, identifying lack of "a mechanism" to facilitate data linkage and overcome the "siloing" of data:

> *What's harder is having that all work together with all the different resources and making it easily accessible. I think, unfortunately, we're still a fair way off from achieving that. (Expert #1)*

Some experts made the point that some jurisdictions (such as South Australia and Western Australia) had better data linkage infrastructure than others. A university-based expert described working closely with government agencies to discuss how the research team wanted to use the data, informing agency decisions to provide approval or not. Introducing legislation produced a cultural shift in attitudes to data sharing:

> *If you want to get a hold of some data, and the usual response was, "Well, why do you want it? Why should I give it you?" to being, "Why not?" (Expert #2)*

One expert reflected that large private companies were undertaking more sophisticated linked data analysis than government or researchers, referring to Facebook, mobile phone companies, Coles and Woolworths. One example linked data on population health data with sales of soft drinks in areas with known rates of obesity. Despite the challenge of accessing privately owned, commercial data, this kind of cross-sectoral data linkage enabled their team to explore interesting questions.

Technical ability and technological infrastructure to undertake data linkage

The ability to match individuals across datasets may still pose a technical hurdle, but experts note that without the investment into building infrastructure and capacity, data linkage cannot progress.

Exact linkage can only occur if unique identifiers have been agreed, then developed, in tandem with appropriate infrastructure. Experts contrasted Australia with countries (such as France and Sweden) where people have social security numbers from birth, thereby facilitating data linkage and tracking over time. Although Australians have Medicare numbers, experts were aware that they couldn't be used for data linkage, because legislation prohibits such usage. Several experts felt that data linkage was easier in countries that had a unitary system of government (unlike Australia's federal system). Several participants mentioned efforts to introduce a national identity card (the Australia Card) in the mid-1980s, which was shelved due to privacy concerns. Another noted that "the privacy lobby is very strong in Australia," although this hasn't prevented some linkages (such as Robodebt). One expert remarked that this has led in part "to the refinement and development of data matching" (Expert #8). Generally, the availability of unique personal identifiers was categorised as a technical issue, rather than one of social justice (related to ethical and political issues) or lack of infrastructure.

Having the technical capacity and technological infrastructure required to analyse big data was described as an "emerging technical challenge" by one researcher (Expert #2) who recounted occasions where computers were not powerful enough to undertake desired analyses. Another expert rejected the notion that privacy laws impeded researchers' efforts to link data, stating that it was "the legacy systems [old and new] which wouldn't talk to each other" that were, in fact, proving to be a barrier (Expert #11). By way of example, this expert spoke about out-of-home care systems and databases that did not accurately track nor link siblings in foster care.

Technical skills and capacity

The idea that "it's not about how much data you have, but rather what you do with it" (Expert #10) came through strongly in several interviews. Experts emphasised technical skills, but also knowledge about data quality, data cleaning, secure data storage, data risks and sensitivities. They highlighted the importance of bringing technical skills and content knowledge together.

One expert (#10) spoke about data linkage as being dependent on the capability of the "data people", who had "the skills and expertise to do things in an appropriate and confidentialised way." Another felt there would be growing demand for people skilled in managing data linkage, with specialised skills around confidentialisation and de-identification, effective and efficient use of large datasets, risks and sensitivities, and proper safeguards and arrangements for data linkage:

> *It's not just a technical or technological dimension where you put in the numbers or you crank the handle and out comes the result. We actually want people to understand what they're doing, understand some of the opportunities and risks and work with the data in a very sensitive way. (Expert #13)*

Another recognised that data mining/exploration required context-specific knowledge in addition to the technical skills, here in the context of assessing a baby's wellness:

> *I think data cleaning still requires the researcher's eye, to know what's important and what's not important. Knowing the content of how a well-baby check is conducted; what do the nurses do in that well-baby check? Well, if you know that, then you know what information ought to be there… I don't think it's a fair ask of technical people. Some balance has to be struck. (Expert #2)*

Another expert described how their organisation used to take agencies' data, analyse it and hand it back, but that they have changed their working model to ensure that the analysis was informed by content knowledge in addition to technical skills:

> *So what we do now, which is the ideal way of working, is that you scope the problem together and each bring your strengths to the project. But then you check in. We have every 2 weeks what they call a showcase. But it's really just a project update. You literally come together to look at what the results are showing and to see how the data structure is, the quality of the data, what's relevant, to actually discuss the project as it's going through. So you come out with a better product at the end. (Expert #3)*

PERCEPTIONS OF DESIGN AND INTENDED USE OF DATA LINKAGE

When asked about projects which used big data or data linkage, examples covered a range of topics, including health, housing, out-of-home care, and school-based funding. Some corresponded to projects where the experts were involved, or had been undertaken by other analysts or researchers. Some of these examples had generated insights that could potentially lead to significant cost savings for government, whilst others were exploratory, seeking improved policy responses, or examining the effects of interventions.

Instead of listing examples by topic, we organise these examples in terms of the different kinds of analyses that they were designed to support. Cross-sectional data provides a "snapshot" of a portion of the population, at a specific point in time, and is often used to address exploratory questions. It may also be used to identify specific cases that could lead to cost savings or other governmental actions. Longitudinal data was particularly useful for evaluating interventions or supporting time-sensitive service or policy responses. Some examples concerned multi-purpose linked datasets with a diffuse or undefined purpose. This contrasted with specific purposes, such as compliance, as occurred with Robodebt. Finally, experts identified several biases arising from the way that the original purpose was defined.

Cross-sectional data linkage

Cross-sectional data was found useful for targeting policies on mental health issues in young people. One expert described how personal attributes from administrative data were linked with primary survey data on specific mental health conditions, commenting that: "I think that's an innovative way to think about things" (Expert #2). Data linkage could also be used to identify sub-populations more appropriately, to examine fair distribution of resources or other social justice issues. For instance, one expert described data linkage for allocating school-based funding. Data linkage that was based on student demographics and parental income from taxation and census data improved fairness compared to previous allocations, which had only considered the socioeconomic status of the area containing the school, although many students reside outside the area. This was considered "a way of more effectively delivering on the Gonski principles and approach" (Expert #9).

Another project examined the use of cladding in Australian buildings, including public housing (prompted by the 2017 Grenfell Tower fire in London) and building approvals. This brought together many government data sources, linked via an external agency, the Data Analytic Centre (DAC). Data linkage helped identify cases that warranted further investigation.

Longitudinal data linkage

Longitudinal data is measured at multiple points in time. Thus linkage of such data requires matching to be repeated at multiple time-points, typically only feasible for a subset of individuals. The key benefit of longitudinal data is that it permits analysis of population trends over time, as well as individual trajectories. Importantly, longitudinal data may permit examination of causal pathways, for instance, comparing characteristics of individuals with and without, and/or before and after, some intervention or treatment. Matching individuals over time requires stronger protections of confidentiality, and also requires longer-term commitment and funding.

Experts provided several examples. Interventions designed to improve social justice were addressed by linking records for youth, comparing their disciplinary profiles at school with their criminal records in adulthood. A retrospective study focused on justice outcomes for youth who were all pre-schoolers at the same time. By linking justice and education datasets, statistical analysis mentioned by one expert (#7) found that First Nations children were "ending up three or four times more often in the youth justice system." These findings were revealing:

> *The structural accumulative risk factors, if you like, that affect First Nations children and their families and their communities multi-generationally, those challenges are so formidable that even the very best long-term, sustained, holistic family support initiative by an agency embedded in the community and in partnership with the schools, still did not shift the dial, this is really important. Yes, it's very disappointing, but we've got to do better, and we know how to do better. (Expert #7)*

This evidence base reportedly helped motivate and justify funding towards a better response.

Experts mentioned three other examples of longitudinal data linkage. These examined factors affecting changes in wellbeing over a time period. In the first study, out-of-home care data was linked with Centrelink data. A study focused on the 5-year period after young people left out-of-home care looked at payments received and movements between different payment levels. Population health and hospital admissions data were linked in a second study examining how a GP's awareness of a patient's diabetes can reduce hospital admissions, and reduce hospital costs by two-thirds. This helped justify increased effort for testing diabetes, and implementation of a call-back scheme from GP clinics to patients who have tested positive for diabetes.

> *So this is a very simple, very doable solution to what is a very expensive problem. Given that over 20% of the population could be diabetic without knowing it .... And given the complication that diabetes can cause, this would be a very cost-effective way to deal with a complex problem. (Expert #12)*

The third example of longitudinal data linkage that was mentioned, a project called "Platforms," deployed at The Royal Children's Hospital, Melbourne, is "a place-based, community-led approach developed by the Centre for Community Child Health that seeks to improve the environments and experiences of children in the communities in which they are born, live, learn and grow" (https://www.rch.org.au/ccch/platforms/). One expert (#4) described the advantages of Platforms' evidence base, constructed via data linkage, as "peer-reviewed, long-term, longitudinal datasets, tracking children and families." These data were used to examine health, employment and other outcomes for children engaged with particular interventions. The access to evaluation data affected policy and funding: "Well, if I'm in government, I want to know that, because I'm going to fund that rather than the [other program]" (Expert #13).

Diffuse or unspecified multiple purposes for data linkage

In contrast to situations where data linkage produces cross-sectional or longitudinal data, data linkage may also be performed for multiple purposes or no specified purpose.

"Their Future Matters" project delivered by the NSW Stronger Communities Investment Unit linked multiple datasets from different agencies, including the Departments of Community and Justice, Family and Community Services, and Health (Taylor-Fry, 2018). Linked data (7 million records from 27 years of data collection across 11 agencies) were collated to support predictive analytics, "providing the unique ability to measure service effectiveness and outcomes for specific sub-populations and characteristics" (Expert #13). This is an example of multiple, diffuse potential purposes.

Robodebt: Purposed for compliance

Two experts raised Robodebt (unprompted) as an example of data linkage, with one of them noting that:

> *The big issue you'll have is the question mark of use of data because you mentioned [de-identified] data and Robodebt; the use of data and data linkage for compliance versus for predictive analytics, and understanding people's pathways through the multiple systems of social welfare and health—and they can be two very different things. (Expert # 10)*

This example highlights the perceived importance of purpose for data linkage.

Biases in defining purpose of linked data

Several participants expressed concern about biases and assumptions underlying the collection and use of big data involved in data linkage. Some concerns were ontological and axiological: data could be shaped to serve particular purposes, with different meanings, but could also be used to fulfil purposes beyond the original intent.

Experts questioned whether big comprehensive datasets were necessarily fit for every purpose. One expert observed that big data analyses can produce biased results, noting that even if the data captures 80% of a population, the missing 20% "is a large enough sample to miss, to actually skew the representativeness of the data" (Expert #9). They reasoned that smaller, well-constructed surveys can produce better quality, more representative data than big datasets.

Other issues centred on how biases may enter into data analytics. Small changes in what data are collected can subtly though significantly alter the meaning of analyses. For instance, one expert (#8) expressed concerns that administrative data "reflects the system, rather than the people." An example was crime data, which is based on who is arrested, rather than those who commit crime, so that some groups (such as Aboriginal and Torres Strait Islander peoples) may be over-represented in arrests in some regions. Another expert expressed concerns about how gendered assumptions may affect data collection and analysis. More generally, analytics that aim to describe, explain or predict some phenomena often link data on outcomes with other data on explanatory factors or predictors. Experts highlighted that researchers select measures of outcomes and predictors, depending on their theoretical viewpoint.

PERCEPTIONS ON DATA LINKAGE MISALIGNED WITH ORIGINAL PURPOSE

During the interviews, experts provided examples of misusing linked data, which could be attributed to: miscommunication of the purpose; or misuse, for a purpose other than that originally communicated. In addition, usage of linked data could also be hampered by: reluctance to share, due to potential misuse; or lack of specificity in linked data, affecting potential use. Discussion of the issues of privacy, data sharing and trust is deferred to the next study (Chapter 5).

Miscommunication of the true purpose

It was acknowledged that it was very difficult for people to provide consent concerning how their personal data would be used. They may often unwittingly share their personal data in the online environment, for example, ticking a box indicating they had read the terms and conditions, or through their use of social media and loyalty cards. Experts noted the need for greater transparency by researchers and data analysts undertaking data linkage, and suggested that they needed to be more vocal about:

> *how they would like to use the data, how they will treat the data properly, the ethical approaches they take towards transparency, integrity, security of information, as well as the benefits to the community from the research they're doing. (Expert #9)*

Use of linked data, for purposes other than originally communicated

Experts were strongly aware that data linkage needs to be targeted to enable analytics for a particular purpose, and that it is unethical to use data beyond that purpose:

> *I think it [linked data] can be used for anything. […] but what you've always got to keep [in mind] is you've got to keep that policy and program purpose. Otherwise, if you're playing with data just for observation, that's when I think it becomes risky. It's not what you want to do. It's the same as doing that with agency data. (Expert #3)*

Another (Expert #11) spoke of international evidence to suggest that the types of programs/interventions that arise from the desire or "paternalistic duty" to solve some wicked problems can "embed bias, prejudice, discrimination and stigmatisation" because such data, ostensibly collected for the social good, is also used to track individuals. Along the same lines, another expert suggested that such targeting of individuals may be driven by cost savings, disregarding the impacts of entrenched poverty.

Experts discussed issues of data being used for purposes other than those used to justify the initial collection of the data. One expert (#11) spoke about Aboriginal and Torres Strait Islander communities not wanting to share their data, but effectively being blackmailed: "They've been told by the authorities that unless you cough up your data, we will cease funding you." This expert spoke about a case where a Native American tribe's genetic data was used without their consent to undermine their land rights claim, by arguing that they did not come from the region their cultural mythology/beliefs said they did. The damaging impacts of this data misuse was one that Aboriginal and Torres Strait Islander groups were reportedly familiar with.

Fear of misuse may underpin reluctance to share

System-level challenges noted by experts often concerned data sharing more generally, with experts referring to reluctance to share data by government departments, including barriers to sharing data across state/federal jurisdictions.

Unwillingness was reportedly driven by two main concerns: agencies were "very worried that somehow we'll use [the data] to punish them, and also due to privacy concerns" (Expert #6). For instance, despite legal provisions, multi-agency sharing of data about individual children in one Australian jurisdiction was still impeded. Another expert spoke about difficulties trying to access school attendance and suspension data, attributing the impasse to concerns that the data would end up "on the front [page] of the Daily Telegraph" (Expert #7). These concerns were echoed by another expert (#9), who noted that some agencies were "very careful and cautious" about who they shared data with "and that might be partly for good reason and partly because they don't want it to be exposed."

At a national level, experts commented on the relative ease of data sharing in countries like New Zealand, which have just one level of government. However, in the Australian context, they noted that data sharing was better within some sectors or agencies, notably health and welfare. For instance, the Australian Institute of Health and Welfare (AIHW) works with states and territories to collect consistent data. Another expert noted a concerted effort on the part of the Commonwealth to work more effectively with the jurisdictions to enable better data linkage. Sometimes data linkage was complicated by division of responsibilities, especially in relation to family support services. When funded at state rather than Commonwealth level, this expert noted that there is no requirement to collect or share information about children.

Lack of specificity in linked data affecting potential use

One expert spoke about difficulties accessing comprehensive service usage data from non-government organisations (NGOs) contracted to deliver government services, making it impossible to determine what services or programs (including type, intensity, duration) are accessed by the public, specifically children or families:

> *We have a really vigorous, strong community sector; there's some amazing things happening in communities. Government have tried to foster this and it's all based on very laudable and good intentions. But … we've got to build the capacity to systematically collect this information, subject to all the legal and ethical limitations, in order to routinely evaluate the collective impact of multiple services on children. We're already investing a lot; we don't even know what we're investing, state and government, state and federal plus whatever's going on in the voluntary sector. That hasn't been quantified, as far as I know. We certainly don't know what the return on that investment is. Are we wasting our money? (Expert #7)*

In other situations, experts noted a misalignment in the timely availability of linked data, compared to desired usage. Some participants noted a lack of "real-time" data, which can create problems, for example:

> *If a child is referred to you and somebody else has been working in the last 2 weeks intensively, you've got to know that. There's no point in just uploading data every 6 months. It's got to be real-time. (Expert #8)*

An additional challenge affecting data linkage relates to whether data is collected about individuals at a scale that is useful for providing or evaluating services. For example, is engagement in a program aggregated as a total attendance for an event, or is individual attendance recorded?

Relevance and quality of linked data to support informed use

In several cases, experts mentioned that the data available did not match their purpose. One expert expressed frustration that what they considered key outcome data (for their research agenda) was not being collected or prioritised (by state education departments or school principals). Another challenge raised was about poor data quality and data gaps, with the expert (#7) commenting that their research team had to "invent a new measure of the social and emotional wellbeing of children."

One expert stressed the importance of determining the quality of the data provided, emphasising that researchers should not accept at face value that the data provided are accurate. They described efforts to develop estimates to help evaluate accuracy of data, which involved regularly sending data backwards and forwards between the research team and agencies. This expert noted instances where their assessments were that the data were 50% incomplete. Understanding the data context was also deemed important: how the dataset was designed, how data were collected and input. For instance, such communication helped refine or create data dictionaries, defining the data being exchanged both from the perspective of the producer and the user.

ROBODEBT AS A CASE STUDY: RECOUNTING THE ISSUES AND CONTROVERSY

The focus of the discussion now shifts to the Robodebt affair. This provides a basis for the next section, which discusses whether experts' views (expressed in the preceding sections) may reinforce or challenge the lessons learned from Robodebt. Here we outline the main features of the issues and controversy. A detailed chronology of the events that contributed to the Robodebt affair from 2015 to 2021 was compiled by Whiteford (2021: Table 1).

At the outset, the government's Online Compliance Intervention system (OCI) was dubbed "Robodebt" by the media (Sutherland, 2018). As it was initially implemented primarily as a cost savings measure, intended to be "tough on welfare" (Carney, 2019), this project's main intent was to calculate and recoup welfare overpayments (Zalnieriute et al., 2019a). It is salient (Mann, 2019) that a stated aim was to recover 1.7 billion Australian dollars (Commonwealth Ombudsman, 2017). As pointed out by Whiteford (2021: 351), this saving was more instrumental in reversing a budget deficit, whilst forming only a small proportion of the total spending on social security.

As noted in the introduction to this book, data linkage and analytics are not new, having been a core component of the social security system for decades. As the federal department responsible for social welfare in Australia, Centrelink and its predecessor agencies have used these techniques to uncover discrepancies in income and to recover debt from welfare recipients who used the service. Prior to 2016, the Departments of Social Security/Services (DSS) and Human Services (DHS) matched data held by the Australian Tax Office (ATO) with earnings declared by Centrelink recipients. If an apparent discrepancy between the two was uncovered, a DSS/DHS worker would manually investigate (Zalnieriute et al., 2019a: 3). As the process was time consuming, staff were resourced to investigate only high-risk discrepancies, where risk was based on analysis using a risk management approach (Carney, 2018: 2).

In 2016, DHS removed human oversight by introducing a new automated process aiming to quickly recover more debt and address a budget deficit (Commonwealth Ombudsman, 2017). The new OCI program impacted approximately 200,000 people in its initial roll-out between November 2016 and March 2017. Then, "because the data-matching and issuing of letters was performed by an automated system, purported debts raised by the OCI came to be known colloquially as 'Robodebt'" (Senate Community Affairs Committee Secretariat, 2017: 1).

The Commonwealth Ombudsman (2017: 5) identified that several major changes to existing processes occurred simultaneously in 2016-2017, which led to the problems created by Robodebt. These steps essentially removed the department's responsibility and human oversight from the process.

- DHS no longer used its information gathering powers to request information directly from third parties, such as employers. This became the customer's responsibility.
- The OCI system automatically sent letters to tell customers about the income discrepancy.
- DHS moved much of the debt management and calculation process online.
- Customers entered their information directly into the OCI system.

This automation had several consequences. A legal issue was that "the factfinding burden for debt that previously rested on the Department was reversed, arguably contrary to the enabling legislation" (Zalnieriute et al., 2019b: 451). This was related to issues with data analytics used to determine the debt, including a lack of transparency and error of logic in the algorithm:

> *... the process was not transparent. Recipients were not given details on exactly how the data-matching process worked. They were not informed how the discrepancy in income, and ultimately the debt, was generated ... income in Australia is reported to the ATO as an annual figure, but to Centrelink as a fortnightly figure. ... However, the letter sent to individuals did not explain that recording variation in income over the year was important to an accurate calculation of welfare entitlements. (Zalnieriute et al., 2019a: 4)*

This logic was later determined to be problematic (Commonwealth Ombudsman, 2017: 1). A key issue for the new system was that the algorithm, used previously by DSS with human oversight, averaged a person's income over a year, and then calculated the debt owed by assuming that the person earned the same amount in each fortnight over that year. It did not account for people who worked casual jobs or irregular hours, for instance, with short windows of higher wages, such as students working between semesters (Galloway, 2017). Alleged offenders were issued with a debt notice. Furthermore, no information was provided on how the debt was calculated. Some were asked to provide income evidence from up to 7 years prior (Zalnieriute et al., 2019a).

These shortcomings were exacerbated by an inadequate process for review. In a rush to roll out the program, the government ignored design standards set out in a 2004 report on smart machine learning systems established by the national Administrative Review Council (Carney, 2019). Many legal problems with Robodebt could have been avoided if the government had adhered to the report's guidelines (Carney, 2019). The problems were also foreseeable (Whiteford, 2021, Section 10: 344), and also judged deliberate, rather than resulting from oversights or errors.

At the time, these analytics were presumed correct, with poor processes for explanation or review. When recipients attempted to challenge a debt, there were reports of long telephone waits with Centrelink staff who were unable to assist with the matter and who ultimately directed the recipients back to the online portal (Commonwealth Ombudsman, 2017). For those who were able to engage with a Centrelink staff member, "their purported debts were often considerably reduced … evidence received by the committee indicated that reductions from thousands of dollars to a small sum or zero were common" (Senate Community Affairs Committee Secretariat, 2017: 3).

These failures of data analytics were ultimately very expensive for the government. On 29 May 2020, the then Minister for Social Services, Stewart Robert, announced that the government would pay back all the money accrued under the Robodebt scheme, a figure amounting to A$721 million. This acknowledged that the scheme was "unlawful" (Turner, 2021), although a class action continued to be pursued (Hayne & Doran, 2020). Finally, the government agreed to settle and pay A$112 million, in lieu of interest, to eligible group members of the class action. In total, the scheme cost A$1.2 billion to undo, excluding the costs of implementing the repayments (Whiteford, 2021: 349).

Robodebt has also been criticised for social reasons because it unfairly targeted welfare recipients who were already "doing it tough." People on welfare and young people were vastly over-represented in those affected by Robodebt (Whiteford, 2021: 348). Recipients included people with disabilities, and people from Culturally And Linguistically Diverse (CALD) backgrounds. Also included were those who may not have access to computers or have enough computer literacy to negotiate the online portal, which had repeatedly been described as difficult to use (Senate Community Affairs Committee Secretariat, 2017). Recipients reported feeling fear on receiving a debt notice, and instead of questioning or challenging the debt, many assumed it must be legitimate because it came from the government (Senate Community Affairs Committee Secretariat, 2017). Some sought the support of their local state Legal Aid service, where recipients and their lawyers invoked the Freedom of Information Act (FOI) in order to get Centrelink to reveal how the debt was calculated (Victoria Legal Aid, 2019, submission 111, cited in Senate Community Affairs Committee Secretariat, 2017: 28). Overall, "The robo-debt initiative actively undermines community mental health" (Victoria Legal Aid, 2019: 19).

Despite criticism and legal challenges, leaked documents from August 2019 revealed that the department was planning to expand the program to target aged pensioners, in an effort to recover a further A$600 million in the budget shortfall (Henriques-Gomes, 2019). However, the government subsequently lost a number of legal challenges, including a class action suit against the scheme (SBS News, 2019)[†] In response to the threat of a class action, in 2020, the government suspended the "averaging" component of Robodebt and announced that, henceforward, each assessment would be overseen by a Centrelink staff member before the letter of claim was sent to the recipient. Leading up to this decision, the government conceded in 2019 that the scheme was "legally insufficient," eventually terminating it in 2021 (Turner, 2021). The aftermath of Robodebt continues, with political and policy responses to redress the way it undermined the integrity of public administration, one of the benefits it professed to address (Whiteford, 2021).

[†] *In addition, grassroots campaigns such as #NotMyDebt helped provide a voice for those affected (see Mann, 2019)..*

ASSIMILATING THE LESSONS FROM ROBODEBT INTO PRACTICE

The Robodebt controversy has progressed significantly since the timing of interviews described above were conducted in 2018-2019, before the class actions had been finalised and the Robodebt system terminated. As noted by critics of the institutional responses (e.g., Carney et al., 2019), there seemed to be some delay in acknowledging and then acting on core issues.

It is, however, important to emphasise that the interview questions were created to support a broader analysis, aiming to methodically identify themes of issues with data linkage. Robodebt was only explicitly raised in interviews by two of the experts. In general, interviewers did not explicitly ask about Robodebt when interviewing the other 12 experts. Thus, where responses by these other 12 experts directly mentioned Robodebt, they are indicative of the pre-eminence of this controversial project in their thinking—about linkage of big data involving administrative data more generally. We also examined whether and how the issues associated with Robodebt were raised by these experts.

In this section, we triangulate the experts' views, with lessons learned from Robodebt, using seven axes. The first considers the problematic role of data linkage in the context of compliance. It is important to be aware that data analytics is not always well differentiated from data linkage, in terms of the second axis of successes and failures. A third lesson is that misuse is related to communication and transparency as well as accountability and legality. The culture around use of big data has evolved more slowly than its technical capabilities. Accordingly, the fourth lesson notes that supply and analysis of linked data is not perceived to be like official statistics, and hence doesn't benefit from these perspectives and traditions. The third and fourth lessons lead to a fifth lesson that accountability must accompany technical efforts, and a sixth, that it is helpful to differentiate the limitations of data infrastructure from those of human infrastructure. The seventh lesson is that social licence is more fruitfully negotiated during data linkage, well before the data is made available.

1. Data linkage for compliance can be problematic

We begin by summarising the points made by the two interviewees who were asked to focus more specifically on issues surrounding Robodebt. These two experts shared their in-depth knowledge of the Robodebt affair, while offering valuable insights into the main motivations for data linkage, under development since the 1970s. Like more recent critics of Robodebt, they highlighted that similar initiatives had previously been motivated and implemented with a rationale of cost saving. They also observed that these initiatives aimed to ensure the integrity of the social security system, whilst also seeking widespread acceptance of this integrity. These aims are consistent with general benefits of automation of governmental decision making, as highlighted by Zalnieriute et al. (2019b: 454), that is, "cost-savings and greater speed …. transparency and accountability, predictability and consistency and equality before the law."

Both experts were concerned that, in the Robodebt situation, there was a gradual erosion of a service orientation, which was replaced by a compliance mindset, where technology had overtaken, and conflicted with, policy. They noted a lack of concern about individuals who were underpaid (or not paid), compared to the very small proportion of those who were overpaid, and also indicated that errors in the automated system can cause the system to fail to meet legal standards. This was later confirmed (Whiteford, 2021: Section 3). In addition, 3 years after these interviews were conducted, it has been argued that for compliance, a social licence is more difficult to attain. A social licence is considered more attainable to support governmental use of analytics which aim to predict, rather than describe (a population), or prescribe actions such as compliance (O'Loughlin & Bukowitz, 2021). In their final comments, one expert looked forward to better government guidance and policy to allay public concern about personal data use.

Many of the other experts (in the core of 10 experts, and the two research team experts) identified benefits related to governmental processes and how they had been streamlined through use of big data. Given that most of these experts were employed in the government sector, this is not surprising. The benefits occurred either at the micro scale of providing service to individuals, or at the broader scales of developing or implementing policy. These benefits focus on the outcomes, and appear to presuppose the technical and data science benefits highlighted by researchers, such as large sample size, the ability to characterise populations that are "hard-to-reach" and ensuring that data is relevant to policy making (Harron et al., 2017). Interestingly, the experts observed that analytics for compliance (via Robodebt and its predecessors) were implemented once technologically feasible.

2. Successes and failures of data analytics may be mis-attributed to data linkage

As noted earlier, the flaws in Robodebt arose from incorrect analysis, which equated averaged annual income with fortnightly income. This adversely affected individuals on irregular incomes, who were often in vulnerable groups. This represents a failure of analytics more so than of data linkage, although data linkage facilitated the analytics. Interestingly, most core experts did not mention analytics when initially asked to reflect on their perceptions of big data linkage and analysis, instead focusing on data. Mentions of analytics appeared when addressing other questions, including one expert who distinguished “smart data” from “big data.” This resonates with recent research showing that the agility of organisations (i.e., “smart” thinking) enables their big data analytics capability to achieve both “incremental” and “radical” innovations (Mikalef et al., 2019). It aligns with projections that: “The future is smart statistics” in an era of big data (Radermacher, 2018: Section 3).

Robodebt provides an example of data analytics “gone wrong,” where the scale of negative impacts had not been envisioned. It is interesting that, beyond the two experts specifically questioned on Robodebt, only 4 of 10 core experts mentioned the Robodebt affair, without prompting.

Instead, when asked for examples of data linkage projects (section on Perceptions of “Big Data” and Data Linkage), experts presented primarily positive views of data linkage and analytics. Most of these examples demonstrated the benefits that could be achieved, whilst sometimes encountering hindrances and delays. In recounting examples of Australian data linkage projects, without always mentioning Robodebt, participants distinguished the use of big data for legal (compliance) purposes affecting individuals from analytics to describe broad patterns and trends across populations. However, when asked for examples of Australian data linkage, some experts nominated negative examples involving misuse of personal data, such as the controversy surrounding My Health Record, and general concerns regarding explicit or implicit access to linked data.

3. Misuse relates to communication, transparency, accountability and legality

When asked specifically about issues encountered with data linkage, experts provided several examples of potential misuse by researchers, whilst admitting a few examples of misuse by governmental agencies. Here, the experts’ Perceptions on Data Linkage Misaligned with Original Purpose may be characterised in five ways (below): unclear true purpose, use beyond agreed purpose, fear of misuse affecting sharing, use constrained by granularity, or relevance of linked data.

The true purpose is not clear. Interestingly, when asked for examples of data linkage, experts provided a few examples with diffuse or unspecified purpose, emphasising their benefits. Conversely, unclear purpose was clearly identified as an issue later in the interview, when asked directly about biases of data linkage. Lack of clear purpose was considered rife online where consent may be given without fully appreciating the extent of that consent. Related to this was a general ethical responsibility assigned to data collectors, regarding transparency, integrity, security and communicated benefits. Experts highlighted that this ethical responsibility needed to extend to the ramifications of the purpose, especially relating to vulnerable groups, such as the complex trade-offs of cost saving and social good (relevant to Robodebt), or “extorting” rather than encouraging data provision (where Aboriginal and Torres Strait Islander people had in one case been asked to provide data in order to receive services). This resonates with the “seven veils” of accountability evaded during the Robodebt affair (Carney, 2019) and responsibilities that rest with analysts to uncover these ramifications, especially on policy (Radermacher, 2018: Section 3.5).

Use of linked data, for purposes other than originally communicated. Most experts emphasised that data linkage ought to be targeted for analysis towards a stipulated purpose, and, moreover, use for other purposes is unethical. However, several noted that applications of linked data to compliance (like Robodebt) may be promoted as supporting the integrity of the data, although the main motivation is cost saving. A more nuanced aspect arose in several examples, where the original purpose was coloured by researcher or agency biases or prejudice. This weakness of data science could be bolstered by referring to a strong tradition in qualitative and mixed (qualitative and quantitative) methods contexts to encourage introspection, clarity and upfront transparency regarding the philosophical and practical positions underlying a research project (Pearce, 2015: 42).

Fear of misuse may underpin reluctance to share data. Experts provided several examples where data custodians were suspicious of how their linked data might be misused. Worldwide, processes are emerging for supplying data under explicit terms of use, supported by legal sanctions and regular training for researchers (Harron et al., 2017: 3). Experts were aware of such initiatives, such as secure data storage. In contrast, Robodebt did not encounter these barriers, in part by evading design standards (Carney, 2019). Experts also noted that responsibilities divided across jurisdictions also complicated access. Robodebt perhaps exploited this complexity by avoiding several layers of oversight and intra-governmental checks (Carney, 2019). In their final comments about future directions, experts focused more on expanded availability or provision of data, rather than the mechanisms that would enable this to occur.

Lack of granularity in linked data affecting potential use. Experts desired data at finer scales to enable more informative data linkage, for example, of individuals rather than numbers attending an event. They also sought more timely provision. This issue of specificity is central to use of highly specific data for compliance (like Robodebt), and more generally, underpins issues of privacy-sharing-trust, the focus of the second analysis (see Chapter 5).

Relevance and quality of linked data to support informed use. The experts' vision for the future focused on unmet desire for greater accessibility and better quality in linked data. In addition to the lack of granularity discussed above, cited examples revealed frustrations with lack of alignment of priorities governing data collection across data custodians (in government) and end users (researchers). Their views reflect that the pathways and relationships related to supply and demand for linked data are far less developed than for well-established official statistics. The latter are built on fundamental principles of impartiality (Struijs et al., 2014: 1), as well as empowering the people, and supporting open and democratic society (Radermacher, 2018: 229). The findings (Commonwealth Ombudsmen, 2017) for Robodebt may be interpreted as a violation of these principles, by what has been described as intentional avoidance of the usual oversight guarding such principles (Whiteford, 2021).

For all five kinds of *Perceptions on Data Linkage Misaligned with Original Purpose*, expressed by these experts, communication and transparency was considered pivotal, regarding the intended or agreed purposes. This also needed to be reflected in legislation, and clarity regarding accountabilities. The juxtaposition of the experts' commentary (3 years ago) with the Robodebt case study reveals several emergent opportunities for overcoming misuse, such as: explicit statements of purpose for data use; scholarly consideration of the researcher position; collaboration and legislation to support data sharing for agreed uses; and greater attention to within-agency accountability, including adherence to good practice. Some issues echoed the general principles of official statistics, including the governmental roles of impartiality, and to empower people whilst supporting democracy.

4. Supply and analysis of linked data is not perceived to be like official statistics

Core concepts underpin many misuse issues (above) raised both by experts and the Robodebt affair. These concern the fundamental meaning of the data and results of analysis. Some misuse could be avoided through admitting the limitations of data used in analyses:

> *In order for official statistics to function as a language for all kinds of societal interactions and decision-making, it is essential that the product 'information' is fit for purpose. This is an essential shift in perspective. (Radermacher, 2018: 225)*

The analysis of big (linked) data spans the fields of data and statistical science, each with subtly different ethos and methodologies. Data science tends to operate on "the data," which is most appropriate for a census or a case study on well-defined populations. In contrast, statistical approaches rely on sampling theory that exploits small, targeted samples (selected via judicious randomisation) to make inferences about the whole population. Users of big data rarely acknowledge that the data, however "big," still represents only a portion of all data on the population: "Typically, Big Data sets are made available to us, rather than designed by us" (Daas et al., 2015). This is understandable, since the technological capability for collating and storing big data emanates from information technology disciplines. In this vein, one well-recognised statistician states that big data linkage and analytics is more like "data engineering" than "data science" (Cressie, 2020), as it is rarely informed by the statistical logic of using data to make inferences or decisions.

One challenge is that linked data is typically neither a census nor is it randomised. It is therefore "composed of observational data and are not deliberately designed for data analysis, and thus [does] not have a well-defined target population, structure and quality. This makes it difficult to apply traditional statistical methods, based on sampling theory" (Struijs et al., 2014: 3). This issue was obliquely recognised by experts. One expert (#5) highlighted that a "better data system" relied on "thinking much more about what other policy questions, policy hypotheses that they're wanting to address." Further, when directly asked about biases, several experts noted several biases (see Biases in Defining Purpose of Linked Data), including overarching concerns about representativeness (and gaps) of big data, which would therefore impact on analysis. Such biases can be addressed by advanced statistical methodology designed to exploit the "bigness" of linked data whilst accounting for the patchy (and unknown) representativeness that is endemic to observational data (Tahamont, 2020). This was echoed by Expert #5, who noted that an ongoing challenge is "how you use large amounts of unstructured data, more effectively." Only one expert noted the methodological challenges of designing and applying quantitative analysis to big data perceived as observational, as a constraint on its meaningful use to support decision making and other official tasks. This is consistent with general calls for more careful application of data science (see, e.g., Efron, 2020; Cressie, 2020), particularly in the official statistics context (Struijs, et al., 2014; Radermacher, 2018).

There was also some ambiguity about the role of official statistics in data linkage and analytics. Here the comments by experts reveal that data linkage was acknowledged as both sitting outside and inside the official statistics realm. Examples of knowledge exchange included the International Population Data Linkage Network (https://www.ipdln.org/), which one expert described as "a useful forum for discussing the technical aspects of data sharing and the challenges." Another spoke of a "very strong level of international collaboration and cooperation, across all of the statistical agencies" in New Zealand (see Chapter 2 above), Canada, the United Kingdom, the Netherlands, Denmark and (to a degree) the United States.

5. Accountability must go hand in hand with technical efforts

Robodebt also exposed fundamental shortcomings in how responsibility was assigned, both within and above a particular government department or agency.

Robodebt was eventually judged to be unlawful, 3 years after these interviews. However, it ''escaped correction'' for 2 years. This delay was attributed by Carney (2019) to four failures related to responsibilities and infrastructure: failures of standards, of due diligence by monitoring bodies, of ethics and compliance by administrators, and of guidance provided by an oversight organisation. At the scale of program oversight, Sutherland (2018: 3) notes that the way that the Robodebt controversy was resolved "provides a good result for the individual applicants but prevents public scrutiny of the program." A notable feature of the Robodebt experience was that its instigation appeared to be a decision, internal to the department, which was made as soon as technical capability and infrastructure made it possible. Critiques have highlighted that the problems associated with Robodebt did not arise from technical challenges per se, but rather, from challenges associated with the social, legal and ethical aspects, many of which were considered avoidable, attributed to failures of oversight (e.g., Carney, 2019; Whiteford, 2021).

Revisiting the experts' commentary regarding misuse (described in the previous section), it is evident that the issues they raised as affecting data linkage also relate to accountability. This accords with the literature. A data linkage program for addressing child welfare (Shroff, 2017: 189) showed that the technical challenges of data linkage and analytics mean that data analysts are de facto presumed to be applying due accountability. However, data science training may have focused solely on the technical elements of data linkage. Thus, this author argues that specialists on the context of data linkage need to be closely involved throughout all phases of linkage and analytics (Shroff, 2017: 191-193). Here, experts echoed these sentiments. When asked about data linkage skills and capacity, several experts noted reliance on technical skills, but also on the "go-betweens" who understood risks and sensitivities of data linkage, and also ensured that these were addressed.

6. Differentiate limitations of data from human infrastructure

In fact, there is widespread potential to confuse technological capability (i.e., that software and analytical algorithms exist and are feasible) with human capability (i.e., skills and knowhow to apply these technologies appropriately). These are clearly differentiated in the business literature when analysing the factors affecting uptake of big data analytics. For instance, Mikalef et al. (2019: 4) identify a data-driven culture and organisational learning as being critical to effective deployment of big data initiatives. Yet these broader issues of a data culture were not always clearly differentiated by our interviewed experts, who often labelled gaps in data available as a technical issue rather than as an infrastructure issue. Nevertheless, Expert #5 notes that data and human capabilities are related: "If we can get the policy thinking right, aligned with the data opportunities, you can get much more granular, much more targeted, tailored understanding of need and issue. But also, potentially how do you actually think about policy intervention."

Amongst the biases noted by experts interviewed here, several concerned a poor conceptualisation of variables and the need to ensure that available big data were fit for purpose. One technical issue they didn't mention was that the false positive rate may be under-reported, as was the case with Robodebt. Some commentary on Robodebt alludes to such technical errors as "procedural" issues (see Carney, 2019: 6), without quantifying the size of the discrepancy, for example, by taking account of the percentage of cases issuing debt notices to those not owing money (Zalnieriute et al., 2019b: 36).

Many big data initiatives exploit new technological capability and are funded to develop the technical infrastructure without also necessarily embedding the human and systems level infrastructure required to support their use. For instance, Robodebt is often described as "technologically-driven." One expert echoed that without high-performance computing infrastructure, it is not possible to perform the data linkage required to address relevant questions. However, the limitations of big data infrastructure systems to house the technology, data and results of data linkage and analytics should not be confused with technological capability (the existence of feasible technology) (Mikalef et al., 2019).

In many cases, failings of infrastructure to support data linkage amount to inability to source data to address particular problems, as highlighted by several responses from experts to interview questions about biases. This underlines the need for long-term data collection strategies, where the questions of the future are used to ensure that the data are collected at the right time. As noted by Struijs et al. (2014: 1, 5), national statistics bodies (and, to some degree, government departments) satisfy the public's need for continual maintenance of "impartial statistical information," which exploits their "privileged access to government data sources" that are not available to business, and, sometimes, researchers. Given that official data is "enshrined in law," this affords a level of trust that does not always apply to other data that relies on voluntary supply. When asked about system-level challenges and service-level challenges, participants expressed many frustrations with data sharing processes that did not permit access to data that existed. This affected requests for data to particular government departments by external parties, or internally, by other departments.

To address such data gaps in a timely way, collaboration between all stakeholders has been identified as a key requirement (Struijs et al., 2014), that can "create a real data culture, becoming more flexible and reactive" (Radermacher, 2018: 228). Such a long view is crucial to develop understanding of cause and effect (e.g., of interventions or treatments), which requires data linkage to construct suitable longitudinal data that either support comparison before and after some event or intervention, or reveal short-term trajectories and longer-term trends (Parrish et al., 2017). This would allow the data being collected to better enable government and the public to answer important high priority questions.

7. Social licence precedes data availability

Experts did allude to the need for a social licence, in several contexts, often in the context of enabling desired analyses by providing access to data. This included a reference to the Gonski principles in education, which guide use of linked data to inform allocation of funding to schools. Also, as detailed in a subsequent investigation (Chapter 5), one expert raised data sovereignty for empowering Aboriginal and Torres Strait Islander people to be involved in deciding how data collected about them would be used. Other examples illustrated that the purpose of data linkage may be misrepresented to data subjects in order to elicit data (e.g., situations where data was extorted from Aboriginal and Torres Strait Islander people), or misrepresented within governmental processes, including with the broader community, in order to attain social licence (with Robodebt advertised as ensuring integrity of the social welfare system).

Internationally, the Robodebt affair appears in the Data Harm Record maintained by the Data Justice Lab in Cardiff, Wales (Redden, 2018). This illustrates that such incidents are visible worldwide, and complement a country's internal efforts to address wrongful use of data linkage and analytics. In this case, the literature notes there may be a "silver lining" to problematic projects (like Robodebt) which has raised the awareness of the general public regarding their role in providing a social licence.

> *… offers a glimmer of hope regarding the central role that grassroots activism and community-led campaigns can play in countering unjust automated technological s ystems with human voices. This specifically relates to how conversations about automated technologies can be community focused and inclusive.* (Mann, 2019: 1)

Similarly, Whiteford (2021: 357) mentions reasons "to be cheerful" about the successful interventions by groups and professions traditionally tasked with questioning the system: community activism, investigative journalism, political scrutiny, legal aid and court systems. Carney (2019: Sections 5 to 7) further identified that social licence requires two-way communication, stressing that the government largely avoided these roles in Robodebt: in ensuring that advocacy is heard, by ensuring that advocacy bodies are adequately independent and resourced; to facilitate professional pro bono and civil society initiatives, and to cultivate a supportive culture of public opinion for the vulnerable. These roles align with the views expressed by the two extra experts consulted in this study, who provided insights into the evolution of Robodebt, especially the change in attitude from welfare as a service to a punitive emphasis on those exploiting welfare.

CONCLUDING REFLECTIONS

The experts interviewed in this study emphasised the benefits of data linkage and the potential of big data. Many also acknowledged the legitimacy of concerns and highlighted the importance of communicating effectively about data integrity, data security and data use. Of interest was the varying degree of uptake, in practice, for seven of the lessons that are beginning to emerge from high-profile data linkage projects, like Robodebt. Other than when discussing Robodebt, only one participant adopted a risk averse stance with respect to data sharing and linkage. They emphasised privacy concerns, data biases and unquestioned assumptions about the objectivity and accuracy of administrative and self-reported personal data. Even the most cautious of remaining experts understood that big data is increasingly likely to be used to target individuals (as for Robodebt), and that with increasingly sophisticated technology and analytic capacity the "velocity" of the data will increase. This was welcomed by these experts, to facilitate ever more finely-grained analysis to be conducted in real time, so that, for instance, interventions could be focused on vulnerable families and individuals when they most need them. Only a few experts reflected emerging broader concerns of the particular challenges posed by data linkage when applied to compliance (Lesson 1). In many cases, the success or failure of data analytics was attributed to data linkage (Lesson 2).

Other than public resistance due to privacy concerns, the key barrier to using big data for human service delivery was identified by many of these experts as organisational resistance to data sharing. This was sometimes attributed to privacy and consent issues, sometimes to legal constraints to data sharing, and sometimes to lack of infrastructure, although the latter was often mis-labelled as a technical constraint (Lesson 6). Practices are currently emerging to overcome such barriers through clearer agreements, accountability and legal controls (Lesson 5). At the time of interviews, organisational culture seemed to play a significant role in determining what data were shared, and with whom. Risk aversion and fears that the analysis would be used to embarrass or challenge the organisation were often suggested as influencing reluctance by agencies to make data available to external parties.

There have now been a series of high-profile incidents which have raised the profile of big data in public debate and attracted significant negative media attention. Internationally, this has mainly focused on social media companies, such as the Cambridge Analytica/Facebook affair. In Australia, these experts confirmed that attention has focused on government datasets, including My Health Record, the online census, Robodebt, and, most recently, the COVID-19 app.

The Robodebt experience highlights a number of key issues in relation to big data. On the one hand, the final outcome indicates that the legal frameworks and social infrastructure around the use of data eventually prevailed, albeit after a long struggle by and on behalf of the victims. However, the Robodebt scandal also highlights the legitimate concerns of individuals and organisations that, once data are collected and linked, the data are open to many kinds of misuse. Experts cited examples of misuse related to poor transparency, communication or accountability. A few highlighted the issues of social licence and difficulties of maintaining data sovereignty, where the subjects of data have some control over their use. These experts were aware of public concerns that governments may try to use data for their own ends, irrespective of the purpose and permissions for which the data were originally collected. Robodebt highlighted the opacity of a simple algorithm used to analyse big data, which was not revealed to the public or alleged offenders: very few people understood precisely how it functioned. Such concerns of explainable and transparent use of algorithms are becoming even more important, worldwide across all contexts where data is important, and where there is increasing uptake and complexity of machine learning and AI algorithms (Lesson 3).

Another issue recognised by these experts was poor interpretation or misuse of analytics, related to avoiding accountability for underlying assumptions. Recent literature emphasises that technical and social issues must be simultaneously addressed, in collaboration, which was a shortcoming of Robodebt. More recent commentaries in the literature highlight the need for big data analytics to be "smarter." Some experts focused on benefits rather than limitations affecting data linked for unspecified or diffuse purposes; some were aware of issues of representativeness inherent in most linked data. Such comments resonate with broader themes in official statistics which highlight that big data is not a census nor a carefully constructed sample, so suffers from the issues inherent in all observational data. This highlights the potential benefits for applying some of the strategies and processes from official statistics into the data linkage and analytics realm (Lesson 4).

Interestingly, all three sources examined—the interviews with experts, the Robodebt experience, and the wider literature on big data analytics—reveal that many issues regarding data linkage are enmeshed. Technical issues are inseparable from social issues. Issues of shared purpose, and subsequent collaboration and vision to justify and supply necessary infrastructure, are intimately related to social licence (Lesson 7). Many experts were mostly concerned with data availability, which they re-badged as a technical issue, rather than one of social licence. The fundamental requirement, that analytics be explainable and transparent, was intertwined with lawfulness. Issues of social justice often also arose in the context of data sharing and availability. Chapter 5 shows that the triumvirate of privacy, trust and sharing emerged as central to these experts' commentaries, while Chapter 6 shows how place-based initiatives provide interesting potential for structuring multi-sectoral initiatives—and could also help limit and manage these entangled issues of linkage.

These kinds of entanglements underline the inherent complexity of data linkage, with all sources advocating that, to avoid adverse outcomes, technical solutions must also consider social impacts, and conversely, social solutions must be conversant with technicalities. Interestingly, the analysis of linked big data is not always perceived (by many experts, nor published studies) through the lens of official statistics (Lesson 4). In the realm of official statistics, a multi-sector, cross-disciplinary, and collaborative approach to data linkage has been advocated to facilitate such exchanges. In turn, this could redefine official statistics:

> *Big Data provides opportunities, but in order to make optimal use of Big Data, a number of challenges have to be addressed. This stimulates increased collaboration between National Statistical Institutes, Big Data holders, businesses and universities. In time, this may lead to a shift in the role of statistical institutes in the provision of high-quality and impartial statistical information to society. (Struijs et al., 2014: 1)*

This need for collaboration was implied in comments by experts on the difficulties of obtaining linked data: across sectors, across states, with reference to international examples and networks.

In conclusion, big data and its linkage are powerful tools which can be used by governments, service providers and researchers for beneficent, benign or harmful purposes. Linking data for service delivery, as opposed to research and policy development, is particularly vexed, since it relies on identifying individuals in the data. In turn, this raises urgent questions of consent, trust and power relations between service providers and service users. Robodebt accentuates that the legislation, policy and efforts towards securing a social licence are struggling to keep up with the technological capabilities of data capture, linkage, and analytics. These interviews with experts highlight that some lessons from Robodebt are still being assimilated. In particular, a more multi-disciplinary and cross-sectoral approach will help foresee impacts on individuals and the governments that serve them.

REFERENCES

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V.I. on behalf of the Precise4Q consortium (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medically Informed Decision Making, 20(1), 310.

Best, N., Dallow, N., & Montague, T. (2020). Prior elicitation. In E. Lessafre, G. Baio, & B. Boulanger (Eds.), Bayesian methods in pharmaceutical research (Chapter 5, pp. 87-110). Florida: CRC Press.

Biddle, N., & Marasinghe, D. (2021). Using census, social security and tax data from the Multi-Agency Data Integration Project (MADIP) to impute the complete Australian income distribution (April 14, 2021). Tax and Transfer Policy Institute, Working Paper 8/2021. The Australian National University Centre for Social Research and Methods.

Bloothooft, G., & Schraagen, M. (2015). Learning name variants from inexact high-confidence matches. In G. Bloothooft, P. Christen, K. Mandemakers, & M. Schraagen M. (Eds.), Population reconstruction (Chapter 4, pp. 61-83). Cham: Springer.

Caley, M. J., O'Leary, R. A., Fisher, R., Low Choy, S., Johnson, S., & Mengersen, K. (2014). What is an expert? A systems perspective on expertise. Ecology and Evolution, 4(3), 231-242.

Carney, T. (2018). The new digital future for welfare: Debts without legal proofs or moral authority? UNSW Law Journal Forum, 1-16. http://www.unswlawjournal.unsw.edu.au/the-forum/

Carney, T. (2019). Robo-debt illegality: The seven veils of failed guarantees of the rule of law? Alternative Law Journal, 44(1), 4-10. https://journals.sagepub.com/doi/full/10.1177/1037969X18815913

Commonwealth Ombudsman (2017). Centrelink's automated debt raising and recovery system. https://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. Social Science Research, 59, 1-12.

Cressie, N. (2020) Comment: When is it data science and when is it data engineering? Journal of the American Statistical Association, 115(530), 660-662.

Daas, P. J., Puts, M. J., Buelens, B., & van den Hurk, P. A. (2015). Big data as a source for official statistics. Journal of Official Statistics, 31(2), 249-262.

Debackere, K. (2016). Let the data speak for themselves: Opportunities and caveats. Journal of Data and Information Science, 1(1), 3-5.

De Mauro, A., Greco, M. & Grimaldi, M. (2016). A formal definition of big data based on its essential features. Library Review, 65(3), 122-135.

Efron, B. (2020). Prediction, estimation, and attribution. International Statistical Review, 88(1), S28-S59.

Elkhodr, M., Mubin, O., Iftikhar, Z., Masood, M., Alsinglawi, B., Shahid, S., & Alnajjar, F. (2021). Technology, privacy, and user opinions of COVID-19 mobile apps for contact tracing: Systematic search and content analysis. Journal of Medical Internet Research, 23(2), e23467.

Galloway, K. (2017). Big data: A case study of disruption and government power. Alternative Law Journal, 42(2), 89-95.

Garrett, P. M., White, J. P., Lewandowsky, S., Kashima, Y., Perfors, A., Little, D. R., . . . Dennis, S. (2021). The acceptability and uptake of smartphone tracking for COVID-19 in Australia. PLoS ONE, 16(1), e0244827.

Garthwaite, P. H., & O'Hagan, A. (2000). Quantifying expert opinion in the UK water industry: An experimental study. Journal of the Royal Statistical Society: Series D, 49(4), 455-477.

Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimaee, M., Barreto, M. L., & Goldstein, H. (2017). Challenges in administrative data linkage for research. Big Data & Society, 4(2), 1-12.

Hayne J., & Doran, M. (2020, May 31). Government to pay back $721m as it scraps Robodebt for Centrelink welfare recipients. https://www.abc.net.au/news/2020-05-29/federal-government-refund-robodebt-scheme-repay-debts/12299410

Henriques-Gomes, L. (2019, August 27). Expanding the Robodebt scheme to pensioners would be disastrous, ACOSS says. The Guardian: Australian Edition. https://www.theguardian.com/australia-news/2019/aug/27/expanding-the-robodebt-scheme-to-pensioners-would-be-disastrous-acoss-says

Herde, J. A., Patton, G. C., Young, J. T., Borschmann, R., & Kinner, S. A. (2020). Adolescent and young adult homelessness during the Covid-19 pandemic: Reflections and opportunities for multi-sectoral responses. Parity, 33(10), 84-86.

Johnston, I. with PHAA (2020). Australia's public health response to COVID-19: What have we done, and where to from here? Australian and New Zealand Journal of Public Health, 44(6), 440-445.

Kaminski, M. E. (2019). The right to explanation, explained. Berkeley Technology Law Journal, 34(1), 189-218.

Leech, B. L. (2002). Asking questions: Techniques for semi-structured interviews. PS: Political science and politics, 35(4), 665-668.

Mann, M. (2019). Social (in)security and social (in)justice: Automation in the Australian welfare system. Global Information Society Watch 2019, Artificial intelligence: Human Rights, Social Justice and Development series (pp. 68-72). Association for Progressive Communications (APC). https://dro.deakin.edu.au/eserv/DU:30143270/mann-socialinsecurity-2019.pdf

McCarthy, R. V., McCarthy, M. M., Ceccucci, W., & Halawi, L. (2019). Introduction to predictive analytics. In: Applying predictive analytics (Chapter 1). Cham: Springer.

McNamara, B., Gubhaju, L., Jorm, L., Preen, D., Jones, J., Joshy, G., & Eades, S. (2018). Exploring factors impacting early childhood health among Aboriginal and Torres Strait Islander families and communities: Protocol for a population-based cohort study using data linkage (the 'Defying the Odds' study). BMJ Open, 8(3), 1-12.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. Information & Management, 57(2), 103169.

OAIC (2014) Guidelines on data matching in Australian Government administration. Office of the Australian Information Commissioner, Australian Government. https://www.oaic.gov.au/privacy/guidance-and-advice/guidelines-on-data-matching-in-australian-government-administration/

O'Loughlin, T., & Bukowitz, R. (2021). A new approach toward social licensing of data analytics in the public sector. Australian Journal of Social Issues, 56, 198-212.

Parrish, J.W., Shanahan, M.E., Schnitzer, P.G., Lanier, P., Daniels, J. L., & Marshall, S .W. (2017). Quantifying sources of bias in longitudinal data linkage studies of child abuse and neglect: Measuring impact of outcome specification, linkage error, and partial cohort follow-up. Injury Epidemiology, 4(1), 1-13.

Pearce, L. (2015). Thinking outside the Q boxes: Further motivating a mixed research perspective. In S. Hesse-Biber & R. B. Johnson (Eds.), The Oxford handbook of multimethod and mixed methods research inquiry (pp. 42-56). Oxford: Oxford University Press.

Price, D. J., Shearer, F. M., Meehan, M. T., McBryde, E., Moss, R., Golding, N., . . . McCaw, J. M. (2020). Early analysis of the Australian COVID-19 epidemic. Elife, 9, e58785.

Radermacher, W. J. (2018). Official statistics in the era of big data opportunities and threats. International Journal of Data Science and Analytics, 6(3), 225-231.

Redden, J. (2018). The harm that data do. Scientific American 319(5), 76. https://www.scientificamerican.com/article/the-harm-that-data-do/

Rose, J., Low-Choy, S., Homel, R., & Katz, I. (forthcoming). Enriching thematic analysis with clustering techniques: A quanti-qual approach for interpreting interviews on data linkage involving big data. In R. Cameron & X. Golenko (Eds.), Handbook of mixed methods research in business and management (Part C: Innovations in mixed methods research designs and data analysis techniques). Cheltenham, UK: Edward Elgar.

SBS News (2019, 17 September). Class action launched against Centrelink's controversial robo-debt system. Special Broadcasting Service. https://www.sbs.com.au/news/class-action-launched-against-centrelink-s-controversial-robo-debt-system

Senate Community Affairs Committee Secretariat (2017). Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative (Chapter 1). Canberra: Parliament of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/SocialWelfareSystem/Report/c01

Shroff, R. (2017). Predictive analytics for city agencies: Lessons from children's services. Big Data, 5(3), 189-196.

Stickley, T., Hui, A., Morgan, J., & Bertram, G. (2007). Experiences and constructions of art: A narrative-discourse analysis. Journal of Psychiatric and Mental Health Nursing, 14(8), 783-790.

Struijs, P., Braaksma, B., & Dass, P. J. H. (2014). Official statistics and big data. Big Data and Society, 1(1), 1-6.

Sutherland, P. (2018). Social security data-matching and 'Robodebts.' ANU College of Law Research Paper, Series 2019, 7. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3360169

Tahamont, S., Jelveh, Z., Chalfin, A., Yan, S., & Hansen, B. (2020). Dude, where's my treatment effect? Errors in administrative data linking and the destruction of statistical power in randomized experiments. Journal of Quantitative Criminology, 37, 715-749.

Taylor-Fry (2018). Forecasting future outcomes: Stronger Communities Investment Unit – 2018 insights report. NSW Government Department of Communities and Justice. https://theirfuturesmatter.nsw.gov.au/__data/assets/pdf_file/0003/673284/Forecasting-Future-Outcomes-Stronger-Communities-Investment-Unit-2018-Insights-Report.pdf

Turner, R. (2021, 11 June). Robodebt condemned as a 'shameful chapter' in withering assessment by federal court judge. ABC News, Australian Broadcasting Commission. https://www.abc.net.au/news/2021-06-11/robodebt-condemned-by-federal-court-judge-as-shameful-chapter/100207674

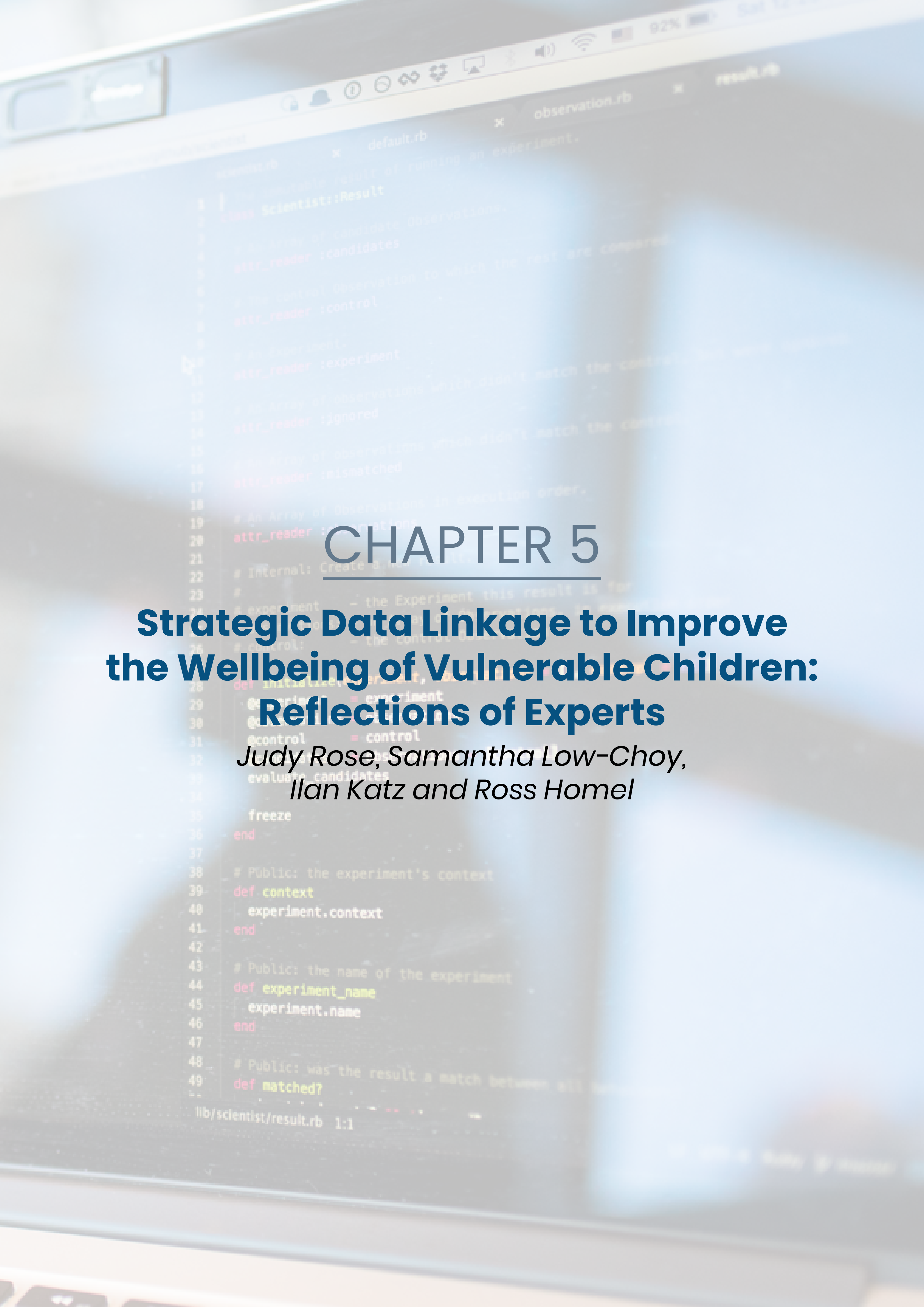
United Nations Statistical Office (1991). Handbook of vital statistics systems and methods. Volume 1: Legal, organizational and technical aspects. Studies in Methods, Series F, No. 35. UN Department of International Economic and Social Affairs: New York. http://unstats.un.org/unsd/publication/SeriesF/SeriesF_35v1E.pdf

Victoria Legal Aid (2019). Intersections between mental health and the legal system and the impacts for people and communities, Submission to the Productivity Commission's Inquiry into the Economic Impact of Mental Ill-Health, April 2019. Melbourne: Victoria Legal Aid. https://www.pc.gov.au/data/assets/pdf_file/0004/241519/sub500-mental-health.pdf

Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco – Or was it something worse? Australian Journal of Public Administration, 80, 340-360.

Zalnieriute, M., Bennett Moses, L., Boughey, J., Burton, L., & Bhathela, N. (2019a). Submission to the OHCHR thematic report to the United Nations General Assembly on digital technology, social protection and human rights (June 4, 2019). UNSW Law Research Paper, Series 2019, 41.

Zalnieriute, M., Bennett Moses, L., & Williams, G. (2019b). The rule of law and government decision-Making. Modern Law Review, 82(3), 425-455.



CHAPTER 5

Strategic Data Linkage to Improve the Wellbeing of Vulnerable Children: Reflections of Experts

Judy Rose, Samantha Low-Choy,
Ilan Katz and Ross Homel

Strategic Data Linkage to Improve the Wellbeing of Vulnerable Children: Reflections of Experts

INTRODUCTION

> *I actually call data linkage the new frontier … Because it's a way in which we can extract more value from existing data. We can collect or compile new statistics, new data insights from existing data without expanding on the respondent burden and without going through difficult data collection costs. I think there's a lot more insight that can be drawn from existing data using data linkage. … We are in this quandary where we're awash with data; they still say we need more data. Which is probably true in some instances, but we also need to use existing data better as well. (Expert #9)*

These insights come from a leader responsible for harnessing big data, at a time when the opportunities and challenges are still evolving. The shifting terrain of this "new frontier" in data linkage is characterised by recent rapid growth in digital technologies and computer capabilities. Consequently, big data have become valuable, both as a resource and a commodity (Fisher & Layman, 2018). The commoditisation of big data is now recognised as an integral part of public and private enterprise (de Souza, 2014); the Australian Information Commissioner reports that "Data held by the Australian Government is a valuable national resource that can yield significant benefits for the Australian people when handled appropriately, and in the public interest" (Falk, 2019: 1). The same forces are at work globally, with the European Strategy Centre referring to big data as, "the lifeblood of the global economy" (European Commission, 2017: 1). A recent World Economic Forum report on Data collaboration for the common good notes that organisations are linking and connecting diverse datasets at an accelerating pace to create value, and this is the key factor influencing the global economy (World Economic Forum, 2019: 8).

However, to realise their value to society, the vast volume and complexity of many datasets necessitate the use of specialised technologies and specific analytical methods and require strict legislative controls and clear policy frameworks that guard the right to privacy and freedom from "dataveillance" (Lupton & Williamson, 2017). An example pertinent to the present study is the construction of "integrated (linked) data assets that would assist policymakers and researchers in relation to children's policy" (Australian Government Department of Education and Training, 2018: 1). In general, it may be the veracity more than the volume of data that matters at the new frontier (Galloway, 2017; Lee, 2017), underpinned by strong safeguards and transparency in terms of decision rules and underlying analytics (Ceri, 2018).

This chapter is the second chapter that is based on interviews with experts who discussed emerging issues and challenges related to big data and data linkage in human service policy and delivery. Here we focus on 10 experts in leadership positions in the education, welfare, health, statistics, and human rights sectors. This "core" group of 10 interviewees, described in Chapter 4, excludes the researchers and those specifically interviewed about Robodebt. This methodology, of eliciting the considered opinions of experts, is well-established in other contexts, but departs from the standard methods used to collate and analyse evidence in this field (as discussed in Chapter 4 under A Note on Research Method). We were especially concerned to better understand the practical, legal, technical, and privacy considerations entailed in the construction, analysis, and use of linked data, given the fledgling status of this field in Australia.

The specific research questions that guided interviews and animated our analyses of experts' views are:

A. What are the benefits and challenges for person and child-centric data linkage?

B. How do the risks to privacy and human rights balance with the potential benefits?

C. In particular, how are child-centric linked datasets used to evaluate the combined impact on child wellbeing afforded by place-based programs and services offered through schools, community agencies, and government departments?

This chapter addresses research questions A and B, whilst chapter 6 addresses research question C. In addition, throughout the interviews, the experts proffered a rich array of insights about many broader aspects of integrated data and data analytics.

In the next section, we provide some context for understanding the information provided by the experts, briefly reviewing a range of policy issues relevant to the construction and use of linked data, particularly data about children. We then describe the project methodology, including selection of experts, interview procedures, and our mixed qualitative-quantitative analysis techniques. We have organised the reporting of our findings around an innovative cluster analysis of the topic or themes in the analyses, allowing us to draw together ideas that were closely related conceptually in the minds of experts, while also according weight to how often topics were mentioned.

The analyses revealed a variety of nested themes in which data-sharing infrastructure (skills, capacity), governance (privacy, security, trust), and human rights (consent and Indigenous data sovereignty) featured prominently. It became clear that these domains are all intimately connected, with reforms in one domain having significant implications for other domains. The capacity for this holistic focus is one of the unique advantages of eliciting and carefully analysing expert understandings of the uses and limitations of data within complex systems.

POLICY CONTEXT FOR LINKED DATA

Developments in data policy

Despite the potential benefits of linking big datasets, and the rapid improvements in enabling technologies, the development of policies on data release, sharing, and integration has been slow in Australia, where this involves government data. This policy vacuum, compounded by jurisdictional differences, exacerbates the default stance of data silos within government agencies and organisations (Productivity Commission, 2017: 121). Overall, data policy development has been:

> *… stymied by a complex, restrictive and conflicting legislative framework that may have been suitable for when it was drafted but is not fit for purpose in a modern data-driven environment. (Productivity Commission, 2017: 175)*

There have, however, been some recent welcome developments. The Productivity Commission report has led to the establishment of the Office of the Australian Information Commissioner and the work of the National Data Advisory Council within the government on the Data Availability and Transparency Bill (Office of the National Data Commissioner, 2020). At the state level, the Queensland Government recently announced an “Open Data Policy Statement,” which pledges that its agencies will anonymise sensitive data to enable data release, whilst upholding the highest standards of security and privacy for individuals (Queensland Department of Science, Information Technology and Innovation, 2016). Other states have taken similar initiatives, including New South Wales, which has established the NSW Data Analytics Centre (NSW Government, 2021). Yet, administrative data collected by government organisations is often associated with varying degrees of informed consent, which, as some scholars note, appears to arise from a move towards a “post-privacy paradigm” (Birchall, 2016: 6).

A recent government scoping study sought to identify key considerations in delivering cross-portfolio insights from linked data assets that would assist policymakers and researchers in relation to children’s policy (Australian Government, Department of Education and Training, 2018). Such data linkages have provided an important resource for identifying risk and protective factors in child maltreatment research (Brownell & Jutte, 2012) and to track vulnerable children from birth (Piotrowska et al., 2020). In turn, this has helped establish early predictors of childhood health and adolescent-onset mental health problems and other adverse outcomes, such as educational underachievement, welfare dependence and criminality (Green et al., 2018). Linking family data also enables tracking of outcomes for previous generations to examine how they affect the health and wellbeing of children in the current generation (McNamara et al., 2018: 5). The option that this data can be linked retrospectively, thereby quickly generating longitudinal data, is a major strength of this approach.

Technicalities of data linkage

Data linkage combines information on the same person from two or more datasets (e.g., surveys or administrative). Records on a person are connected, ideally, through a unique identifier (e.g., name and date of birth, tax file number). Where identifiers are not unique (especially names and addresses), different methods are possible, such as statistical linkage or probabilistic methods (Queensland Health, 2020). For example, information collected on an individual across multiple health-related databases and organisations (e.g., hospitals, disease registers) can be collated for analysis (Brown et al., 2017).

Linking highly sensitive data is managed through carefully constructed infrastructure: different repository organisations may hold different parts of the data (linkage centre, research centre); a "data trustee" may act as an intermediary (Brownell & Jutte, 2012: 121); or data access may be controlled, for instance, via secure sites for in-situ data analysis (Križan et al., 2015). Risks associated with data linkage need to be well managed, to ensure that the privacy of individuals is maintained when integrating or analysing datasets such as health records (Brown et al., 2017), educational records (Department of Education, Skills and Employment, 2021), and other personal information (Productivity Commission, 2017).

Identity in linked data

Data linkage is only possible because administrative datasets can identify individuals. This is most accurate when individuals can be uniquely identified, although a high degree of accuracy may be obtained using probabilistic methods. However, once data are linked, comprehensive protections are essential to preserve privacy (Dankar et al., 2012). Privacy may be compromised when data are collected for case management purposes then later repurposed for quality assurance, research or evaluation. Data suppliers are responsible for avoiding situations where the subjects of data are unaware of how their data are used or are not able to provide informed consent on its various uses (Productivity Commission, 2017).

Protections ensuring security, confidentiality, and privacy of such data may entail sophisticated methods to ensure that no identifiable information is produced in data outputs (including the re-identification of individuals), for example, as used by the Social Analytics Laboratory (SAL) at Griffith University (www.griffith.edu.au/criminology-institute/our-institute/social-analytics-lab). These methods include controls around disclosure that verify that an individual cannot be re-identified from the outputs (Office of the National Data Commissioner, 2020). Some Australian data agencies have adopted the internationally recognised Five Safes Model of Managing Risk (see Figure 5.1) for data custodianship and management (Desai et al., 2016). The Five Safes were designed to limit the risk of disclosure and breaches of privacy, trust, and sharing protocols, but the current challenge is to operationalise these principles.

Figure 5.1. Privacy, trust and sharing mapped to the Five Safes Model

Privacy	Safe projects: appropriate purpose? Safe data: sufficient protections? Safe outputs: results non-disclosive?
Trust	Safe people: researchers authorised?
Sharing	Safe settings: prevent unauthorised use?

Privacy

Privacy is defined by the Productivity Commission (2017: 66) as the safeguarding of personal information that could reasonably identify an individual. This definition is core to The Privacy Act 1988, established to protect the privacy of Australian citizens and regulate how the government manages their personal information (Productivity Commission, 2017: 184). The Act requires that personal information be collected with consent, de-identified, and protected from misuse, including for reasons not intended or by unauthorised entities. Thus, the draft Data Sharing and Release Act (Department of the Prime Minister and Cabinet, 2019) recommends greater efforts to de-identify data and obtain individual consent explicitly for data linkage, where "reasonable and appropriate" (Falk, 2019). Responsible data linkage should also provide guarantees, particularly for vulnerable groups, about what data is collected about them, how it is used, and why it is deemed necessary (Galloway, 2017).

Trust

Trust is involved in every relationship affected by data linkage among data providers, custodians, and the subjects of the data (Brown et al., 2017). Trust of data providers is reflected in the Five Safes Model, which asks: Do people trust data providers to use their data appropriately? Research has found Aboriginal and Torres Strait Islander peoples are significantly less likely than other groups to trust government departments with their data (Biddle et al., 2020). Additionally, trust amongst data providers in different organisations is required before releasing and linking information from data silos in longitudinal research, for example, tracking health outcomes of vulnerable populations (Leonard et al., 2018).

Trust is closely linked to privacy, which may be compromised by poor data management practices (Carney, 2019). Public trust has been tested after the failure of big data systems, such as the online delivery of the Australian Census in 2016. Reputation harmed in this way may affect future efforts (Productivity Commission, 2017), although a relatively high level of trust in the Australian Bureau of Statistics has since been restored (Biddle et al., 2020). Trust also aligns with Safe people: Can the people handling and analysing the data be trusted to use it appropriately, and are they authorised to handle the data? (Desai et al., 2016: 5). Data custodians are ethically obliged to maintain confidentiality and require a "dedicated ethics adviser" to advise on protections for data sharing and linkage applications (Productivity Commission, 2017: 248). Different repository organisations may hold different parts of the data and appoint a "data trustee" as an intermediary (Brownell & Jutte, 2012: 121). Hence, improving community-wide trust on the use of personal data is one of the main objectives of the Data Availability and Transparency Bill (Office of the National Data Commissioner, 2020). Another domain of trust is trust in the data themselves—their completeness, reliability and validity, and potential biases.

Sharing

Data sharing has been defined as making data available to an approved person, agency or organisation under agreed conditions (Department of the Prime Minister and Cabinet, 2019: 4). Thus, data is provided to trusted users under safe and secure conditions that control how data is used, including via data-sharing agreements (Productivity Commission, 2017: 183). Data-sharing applies to Safe settings: Does the access environment prevent unauthorised use? (Desai et al., 2016: 5). Uncertainty about privacy requirements has been perceived to lead to "bad experiences," with potentially negative, long-lasting effects on data sharing and release (Productivity Commission, 2017: 11).

METHODOLOGY

Interviewing experts

Two interviewers from the research team conducted in-depth, semi-structured interviews with experts (averaging an hour in duration) between October 2018 and August 2019. The 10 experts in this study (the "core group" referred to in Chapter 4) include three senior managers, two senior executives, three directors, one chairperson and one senior academic. These experts had decades of experience and worked across a broad range of national agencies, state government departments, public universities, and an intergovernmental body. All experts held leadership positions and were knowledgeable about implementation of data linkage involving big data and administrative data, having a mix of technical, research and policy experience. Hence, they were deemed experts having "comprehensive and authoritative knowledge in a particular area not possessed by most people" (Caley et al., 2014: 232).

Experts were recruited via the research team's contacts, and satisfied criteria of: (1) holding leadership or influential positions in their organisation; and (2) having operational knowledge, policy expertise, technical or research experience with data linkage involving big data. 9 out of 10 experts were based in Australia, working in the states of Queensland, New South Wales, Victoria, South Australia and the Australian Capital Territory.

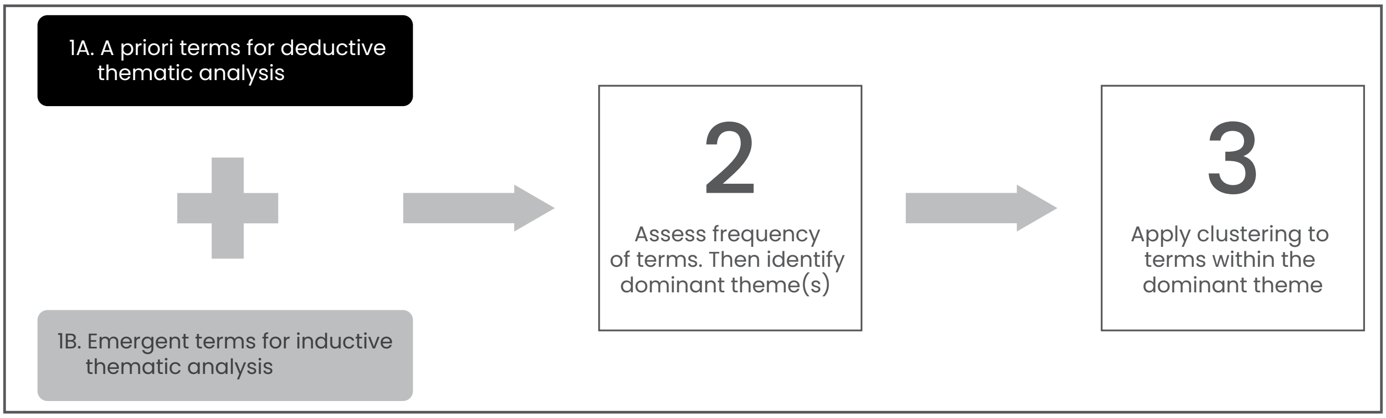
Recruitment of 10 experts exceeds the minimum of 6 and falls just below a ceiling of 12 recommended in expert elicitation protocols (Knol et al., 2010: 6). Since the methodology used to select experts was targeted and purposive, it is not intended to enable statistical generalisation from our sample to the larger pool of perhaps 50 leaders on data linkage in Australia. Constrained by the tension between confidentiality and credibility, we note that the project team acted as “supra-experts”: recruiting experts deemed credible (Caley et al., 2014: 232); and seeking balanced representation across sectors and geographic regions (Knol et al., 2010: 7). In the results presented below, quotations are provided to provide “thick detail” to contextualise expert observations (Polit & Beck, 2010: 1456).

Most interviews were conducted via telephone, a couple face-to-face, and one via video conference. Prior to the interview, experts read and signed information and consent forms. All participants agreed to have their interviews digitally recorded and chose to review questions prior to the interview. Recordings were transcribed verbatim and de-identified prior to analysis. Interview questions adopted an open-ended format to elicit the widest possible range of opinions. All questions underwent a series of refinements by the research team and were reviewed for face and construct validity before full delivery to the 10 experts interviewed.[†]

Mixed methods approach

Our analytic approach mixes a qualitative analysis (see 1A, 1B in Figure 5.2) with a quantitative analysis (steps 2 and 3) to enumerate themes and to group concepts raised in interviews. Applying quantitative methods to data used for qualitative analysis make this a crossover mixed analysis that provides “a higher level of integration” (Onwuegbuzie & Hitchcock, 2015: 284).

Figure 5.2. Strategy for mixed qualitative (black and grey) and quantitative (clear) methods



Qualitative coding of themes

Software (NVivo: QSR, 2018) was used to support researcher-specified manual coding of textual data into themes (Braun et al., 2019). The coding frame was assembled by identifying spoken text segments that were coded to a theme (at the level of paragraphs rather than sentences) so that coded material could be interpreted within the context of the whole interview conversation. Analysis involved systematically reducing data to codes, then developing themes from these codes, to develop a taxonomy of themes arranged in a visual and hierarchical structure (Onwuegbuzie & Hitchcock, 2015, Table 16.1). Each theme was classified as a priori or emergent depending on whether it was derived deductively, through reference to interview questions, or inductively, through reference to interviewee responses (Henry et al., 2015). This process is illustrated in Figure 5.2 (steps 1A, 1B).[§]

Mixing-in quantitative method

The themes selected for clustering included both dominant topics raised by experts and important topics identified in the literature. Segments of text were selected as units of analysis for hierarchical agglomerative clustering via complete linkage (Friedman et al., 2009), as implemented in NVivo (QSR,

---

[‡] *The full set of interview questions and details of the design rationale are available in Rose, J., Low-Choy, S., Homel, R., & Katz, I. (forthcoming). Enriching thematic analysis with clustering techniques: A quanti-qual approach for interpreting interviews on data linkage involving big data. In R. Cameron & X. Golenko (Eds.), Handbook of mixed methods research in business and management (Part C: Innovations in mixed methods research designs and data analysis techniques). Cheltenham, UK: Edward Elgar.*

[§] *All themes are tabulated in Rose et al. (forthcoming).*

2017: About cluster analysis). The presence/absence of codes within each text segment was tabulated internally within NVivo, reflecting either a conceptual connection (where codes are correlated when their meaning is closely connected) or co-occurrence (where the expert mentioned them in the same paragraph). Given this binary input to clustering, Sorensen's similarity index was the most suitable (Choi et al., 2010), from three available in NVivo (QSR, 2017: Selecting a similarity metric). We sought to produce a taxonomy, mapping relationships among themes, rather than identifying distinct clusters. Thus, hierarchical clustering allowed us to examine whether compound themes (such as "privacy, trust and sharing") were more appropriately considered holistically or divided. Sensitivity analysis was used to assess the stability of clusters: did they stay the same when interviews with two research team members were incorporated?

Quantitative analysis is applied in a descriptive way; we do not attempt inferential statistics with this small sample. Quantitative results should be interpreted as exploratory analyses, suggesting patterns that can be tested in future studies. These descriptive statistics are appropriate even with small samples (Tukey, 1977).

RESULTS

The results follow the strategy and logic presented in Figure 5.3.

Step 1A Thematic analysis: research question A

The deductively coded themes mainly contributed to addressing the first research question, which sought to understand the benefits and challenges or barriers to data linkage.

Benefits of linkage

Experts listed benefits ranging from calculating how many children have received assistance and for what benefit and cost (Expert #9), to improving the quality of data on children (through linking multiple sources such as Child and Family Health with Child Protection data), thereby gaining novel insights only possible through data linkage (Expert #1 and Expert #2). Other experts noted that linking data could help to improve the efficiency of services to children (e.g., addressing the duplication of organisations funded for the same purposes) or to provide more targeted care to vulnerable children (Expert #3 and Expert #4). Further, one expert argued that strategic use of datasets, to scale down and link small parts of big data, can reveal valuable insights, or "true gold" (Segment #9a, Expert #10*).

Barriers to linkage

A barrier to linking data is the weight that some decision makers accord to the privacy of people, particularly vulnerable children, regarding their identity and rights to protection from harm. In terms of linking government data, one expert felt there was not enough being done to protect privacy, and that the public need to weigh up the benefits and risks of sharing sensitive data (Segment #4, Expert #12*). Some experts noted that breaches of privacy or mistrust between parties may aggravate the sharing and linking of data:

> *I think, in general, we need to get better at sharing information and data, because we all have a role to play in terms of securing outcomes for children. I can't underscore how difficult the system issues are. Also, there's still a culture of mistrust: "I don't want to share this with you because you might hit me over the head with it." We've got to move beyond those things and find a way to make this work better. (Segment #5, Expert #6*)*

Experts highlighted the need to build trust within the local environment through collaboration and consultation with community residents and agencies who hold datasets, particularly when linking data on vulnerable children and families (Segment #2, Expert #2*). Building trust across these stakeholders is viewed as the major challenge by most experts.

¨ Key illustrative examples of spoken text segments from experts are included in this section, with other segments starred, available on request from authors. Affiliation details are provided under Contributors.*

Related to trust is the concern about data sovereignty:

> *That word "sovereignty" reflects how the data that describes them … The Indigenous Data Sovereignty group is about: this is our data and you're always coming at us from the angle … that we are a problem, that we are a risk … You need to be aware that some of the barriers to sharing of data come back to culture, and people think, "Well, I don't want to give that over because I'll lose control of it." (Segment #6, Expert #11)*

This expert pointed out that Aboriginal and Torres Strait Islander data subjects are seeking control not only of what data are linked, but also how data are analysed and how findings are disseminated. This is strongly related to the issue of consent; what data subjects consent to regarding linkage of their data is related to control of data (Kukutai & Taylor, 2016; Walter, 2018: 261).

Another barrier to successful data linkage relates to the technical skills required and the need for investment in an ongoing program of training and staff development (Segment 20, Expert #1*). The data analysis team requires expertise in the technical aspects of the data and data linkage, as well as deep knowledge of the policy and practice implications of the analysis, illustrated by a recent meta-analysis (Low-Choy et al., 2021): the impact of big data analytics on business performance was moderated by capability (including personnel, skills, and training).

Step 1B Thematic analysis: research question B

The next set of inductively coded themes address the second research question, which sought to understand how risks to privacy and human rights weigh up against the potential benefits gained by linking datasets. One expert felt that greater advocacy for data linkage was needed in the community and among researchers, and linked to the need for "establishing a social licence for effective use of data," and that this would require "bringing the community along" (Segment 10a, Expert #12*). This sentiment was expressed differently by Expert #9, by contrasting social licence against the speed of technological advances:

> *The whole dimension of balancing use with social licence is probably the whole thing that's emerging in our mind where technology has taken a couple of steps ahead. But do we have community support to unleash the power of this data, knowing the potential risks? (Segment #10b, Expert #9)*

Another expert expressed the need for appropriate legislation, legal protections and good governance:

> *The things that we have to take into consideration are the legal framework, the data-sharing frameworks, the social licence … We do have a data-sharing agreement, but it's a pretty light one. What it does allow us to do is to get the data from agencies. If you look at it, makes it easier for agencies to give the information to us. But this is not where it stops. It's very much around the actual projects and the outputs and the infrastructure that sits around it. So, the governance. At the moment, we do the governance per project, as opposed to a single governance framework. (Segment #11, Expert #3)*

Related to good governance is the need for staff skills and training to ensure that consent, confidentiality, and security protections are applied appropriately in data linkage processes (Segment #12, Expert #10*). Such practices would be routine for a "policy and political community that understands both the opportunities and limitations of citizen-based analytics" (Gluckman, 2017: 24).

Step 2: Theme frequency analysis

A frequency analysis was applied to the themes derived from spoken text segments coded for the core set of 10 experts. Altogether, 726 spoken text segments were manually coded into 66 themes. The most frequently mentioned theme we identified as "trust, privacy and sharing," which attracted 34 mentions across the 10 experts (Table 5.1). Table 5.1 illustrates an excerpt of our coding index that shows the relationships between themes, and subthemes that connect to challenges of data linkage. We found evidence consistent with thematic saturation with 10 interviews by comparing the analysis with 12 interviews (Rose et al., forthcoming). This result is consistent with general findings (Brod et al., 2009) that, typically, saturation of topics may occur within 12 interviews.

Table 5.1. Main themes and sub-themes related to challenges of data linkage by number of experts and spoken text segment counts

Theme	Subtheme	Number of Experts (n=10)	Number of Spoken Text Segments
Challenges of data linkage	Individual identity tracking and confidentialit	9	16
	Challenges linking admin and child wellbeing data	6	12
Ethical issues	Trust, privacy and sharing	10	34
	Safety and security concerns and protections	8	15
Practical aspects	Data sharing legislation and frameworks	5	15
	Communicating data analysis and context	5	6
Resource issues	Capacity issues	6	12
Technical issues	Highly technical or data-oriented aspects of data linkage	7	14
Linking administrative data in place-based contexts	Evaluating place-based programs and outcomes	8	19
	Improving service delivery	7	13
	Linking with child wellbeing assessment data	9	11
	Providing interventions to individual children	6	9

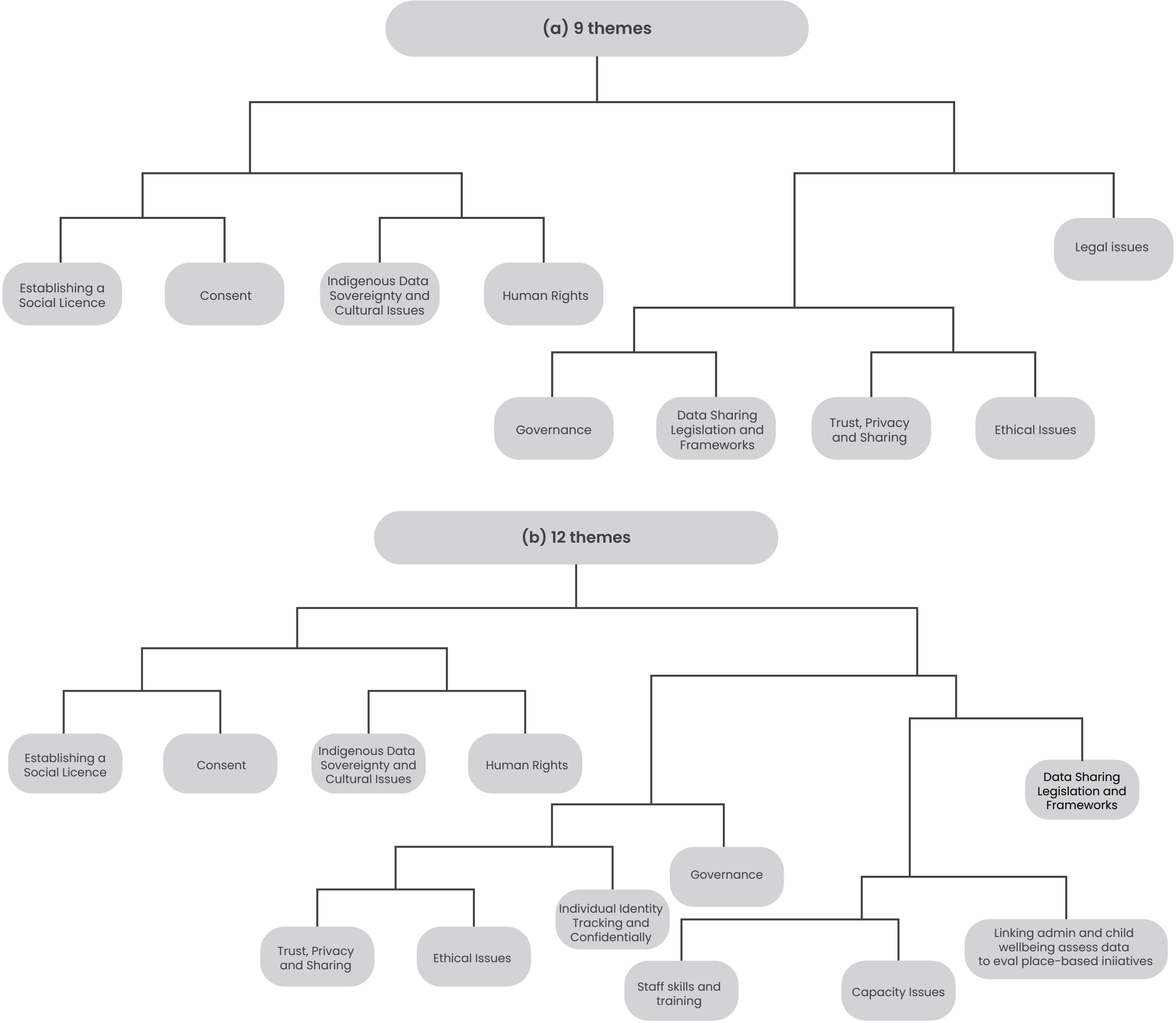
Step 3 - Cluster analysis

Cluster analysis narrowed focus to those parts of the interviews that were associated with the most common theme of “trust, privacy and sharing.” Specifically, we identified nine themes that were cross-coded with trust-privacy-sharing (QSR, 2017: Use coding stripes to explore coding). Cluster analysis was used primarily as a descriptive technique, grouping together themes that tend to co-occur, being mentioned in the same part of conversations. Thus, we used spoken text segments as the units of analysis, rather than experts. Our first cluster analysis (Figure 5.3a) included nine conceptually related themes. This selection of nine themes was validated by a common rule of thumb (which applies to general topics with large pools of participants) that a cluster analysis using nine variables requires at least 2^9 or 512 segments (Dolnicar, 2002: 2). This nine-theme analysis revealed four distinct pairs of related concepts, with ethical issues co-occurring with trust, privacy and sharing, and data-sharing legislation and frameworks co-occurring with mentions of governance. Both of these pairings were conceptually connected to the overarching theme of legal issues.

The next pairing of themes (bottom, Figure 5.3a) show human rights co-occurring with Indigenous data sovereignty and cultural issues, and consent co-occurring with establishing a social licence. Consultation and consent are particularly salient for Aboriginal and Torres Strait Islander peoples who have previous experience with data representing them causing more harm than good (Kukutai & Taylor, 2016). An overarching goal is to establish a social licence that allows people to trust that their data will be used as they have negotiated, and for the good of their community.

To further assess stability of clusters, clustering was repeated with 12 themes (Figure 5.3b). However, those results should be interpreted with caution, since 12 themes exceeds the rule of thumb for sample size. The 12-theme cluster analysis (Figure 5.3b), revealed three distinct sets of ideas. The first set (top, Figure 5.3b) was related to the challenges of linking administrative and child wellbeing data to evaluate place-based interventions. Experts noted that this was limited by capacity issues on the ground (e.g., among community organisations) and a lack of staff skills and training to facilitate or undertake linkage effectively. The overarching concern connecting these clusters was that data legislation and frameworks concerning best practice in linking data on services and place-based initiatives for vulnerable children and families are still being developed. (This complex issue is developed further in Chapter 6.)

Figure 5.3. Simplified dendrogram from cluster analysis of key themes using (a) 9 themes or (b) 12 themes



A second grouping of concerns (middle branch, Figure 5.3b) comprised ethical issues associated with securing privacy (in respect to identity and confidentiality) and establishing trusting relationships at all levels: with families and children, with community organisations that service children, and with data custodians. This concern was connected to the overarching need for data governance. Maintaining public trust relies on robust data governance (sound management decisions concerning the availability, usability, security, and privacy of data), as well as clear and transparent processes around data usage (Lee, 2017).

A third cluster of concerns (lower branch, Figure 5.3b) revolved around human rights in data linkage, in particular, ensuring the right of children to be protected from harm and managing the risk of stigma and bias (Eubanks, 2018). For example, an unintended consequence of the release of public data can lead to the naming and shaming of schools or communities that vulnerable children attend (Rose et al., 2020). Finally, in this third cluster (b), human rights, once again connected conceptually to issues of consent and Indigenous data sovereignty.

DISCUSSION

Our overall finding, encapsulated most clearly in Figure 5.3b depicting the clustering of 12 key themes, is that experts discerned an interrelated set of themes in relation to data linkage:

a. Data-sharing legislation and infrastructure that build the resources, skills, and capacity for linkage;

b. Governance of data linkage to address ethical issues and privacy, security, and interagency and public trust;

c. Human rights concerns, including Indigenous data sovereignty; and

d. Gaining consent from vulnerable groups to link data and establishing a social licence through collaborative and trusting relationships between data custodians, analysts, and data subjects.

None of these themes can be accorded more weight than any other; decisions made in one domain will have important effects on other domains. For example, gaining consent and building trust would be much easier with legislation that has the flexibility to allow highly trained staff to negotiate with data custodians and data subjects at the local level in order to strike an appropriate balance between the right to privacy and the public good. Such negotiations presuppose highly developed governance arrangements in a data linkage agency, and the technical capacity to deliver high-quality linked and integrated data in a variety of forms with a range of safeguards for different stakeholders. Accordingly, legislative and policy reforms must proceed on all fronts simultaneously, ideally within a framework that facilitates continuous quality improvement in light of experience.

Benefits

In this chapter, experts highlighted that data linkage for child protection would help to better target and evaluate services. This aligns with existing recommendations (e.g., Cuccaro-Alamin et al., 2017) and evidence (e.g., Vaithianathan et al., 2013), but contrasts with other findings that the potential benefits of predictive analytics in child protection are outweighed by new problems that are introduced, including hidden bias in the algorithms, the incompleteness and unreliability of the datasets, the lack of transparency, and the impact upon families (Munro, 2019). Expert views were consistent with a review of data linkage in the health sector by Holman and colleagues (2008), which demonstrated that data linkage has helped improve the evaluation of the appropriateness of how these resources are allocated, whilst also casting light on the effectiveness, efficiency, and equity of initiatives.

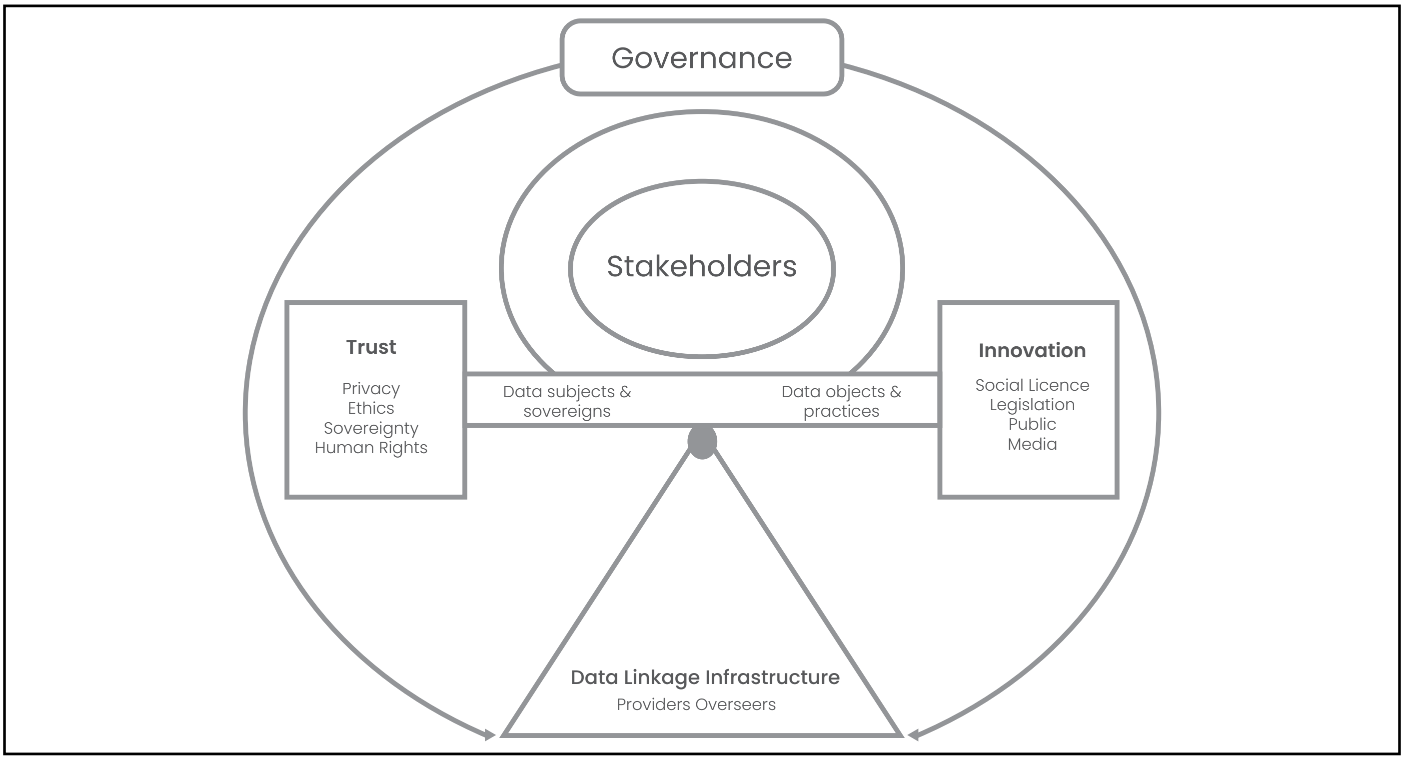
Trust

Experts' concerns about trust resonate with research that has found that, in Australia, public trust in organisations handling personal data is low (Biddle et al., 2020). For those experts in leadership positions responsible for utilising or providing linked data, the negative issues tended to be more focused on the mechanisms for sharing data, as reflected by the themes of consent, governance, organisational barriers and jurisdictional differences, resources, and costs.

Two interesting findings were related to a compound theme, on privacy/trust/sharing, which emerged from the inductive thematic analysis. Cluster analysis confirmed it was closely linked to legal, ethical and governance issues. This triple-barrelled theme is not presented as a compound construct in the literature on data linkage, although previous authors have stressed the potential to balance privacy, trust, and data-sharing considerations. For instance, to balance these three issues, Holman and colleagues (2008) have argued the feasibility of using independent organisations for data linkage in the health sector to promote community development and improve community outcomes, whilst also conserving individual privacy.

We can compare these findings to a recent typology of trust applying to linkage of administrative data (Sexton et al., 2017; Figure 1), which identifies a foundation (triangle, Figure 5.4) of infrastructure, comprising trusted third parties, secure setting providers, government data providers and data oversight bodies, servicing a broad range of stakeholders (circle, Figure 5.4). The fulcrum balances issues that affect or pose risk to data subjects and sovereigns (left, Figure 5.4) with issues that apply to data as objects and practices (right, Figure 5.4). The first apply to individuals and include privacy, trust and ethical issues such as data sovereignty and human rights. Issues also apply to data as an object (including data linkage practices by data bodies, data custodians and data users) for the purposes of innovation (e.g., improving the wellbeing of vulnerable children). Innovation for a social good requires government to establish legislative frameworks through which a social licence can be forged with the public and other fora (e.g., media). To achieve this good governance is required (encircling arrows).

Figure 5.4. Extending conceptual framework by Sexton et al. (2017) to incorporate a balance of trust and innovation



Data sovereignty

This theme emerged during an interview with a single expert. It directly relates to a major grievance of Australia's Aboriginal and Torres Strait Islander peoples regarding the "accepted" ways of aggregating data to highlight the "5Ds" of difference, disparity, disadvantage, dysfunction, and deprivation (Walter, 2018: 258). Measures of these constructs are typically detached from Indigenous cultural contexts and the legacy of colonisation. As noted by Kukutai and Taylor, this is a complex problem, since:

> *The multifaceted nature of indigenous data sovereignty gives rise toa wide-ranging set of issues, from legal and ethical dimensions around data storage, ownership, access and consent, to intellectual property rights and practical considerations about how data are used in the context of research, policy and practice.*
> *(Kukutai & Taylor, 2016: 2)*

An important facet of sovereignty is Indigenous involvement in governance arrangements around the collection and management of data, since typically, data practices have not served Indigenous people's interests or improved the lives of vulnerable children (Kukutai & Walter, 2015). However, the Indigenous Data Sovereignty movement has started to reframe issues of trust and data sharing in ways that allow First Nations people to have more control in setting the data agenda rather than being positioned as the subjects of data (Walter, 2018; Walter & Suina, 2019).

Technical issues

Some technical issues are so pervasive in the human services sector that they can be easily overlooked. These include data whose quality on inspection falls far short of the claims of management and the expectations of evaluators, as well as, in some cases, the total absence of data at a level that would allow the activities of an agency to be linked in any meaningful way with outcome measures (Homel et al., 2021).

There are other technical issues that arise during data linkage that are not always apparent to data producers or end users. Some were not mentioned by our sample of experts, despite their technical expertise and experience that was evident in their discussion of technical issues in general. A good example is biases in linked data. Biases may arise from how individuals have been selected or not into the dataset, for example, via self-selection (e.g., Tam & Kim, 2018) or the lack of representativeness of the population of interest (Cox et al., 2018), or from discordant harmonisation of data from multiple sources.

Bias can also be created by analytic procedures, an issue that has attracted substantial attention in the literature. The way that big data is used to automate decision making has been under question for at least 5 decades from various perspectives: computer science (Quarteroni, 2018), mathematical statistics (e.g., Cox et al., 2018), official statistics (Tam & Kim, 2018), and social science (e.g., in policing, Brantingham, 2017). The United States has adopted a relatively unregulated approach towards uses of big data in contrast to the European Union's decision to establish clear protocols (Kaminski, 2019) which aim to ensure that automated algorithms are "explainable" (Selbst et al., 2019). The latter address the inherent biases that can lurk within any dataset, regardless of its size, with big data often associated with over-confidence (Cox et al., 2018). To compensate, a multi-method approach to analytics has been advocated, based on a deep knowledge of theory and statistical methods (Ceri, 2018).

Other perspectives

Despite their technical expertise and the many comments they made about technical challenges, perhaps it is fair to say that, in comparison with some other overviews of the field of data analytics, our experts put relatively less emphasis on technical issues, or service delivery issues at the individual level (Hopf et al., 2014), and relatively more on policy challenges. This undoubtedly reflects the seniority of many of our interviewees and their long years of experience in attempting to use linked data across government for policy purposes. A good example of the primacy of technical considerations is Gluckman's (2017: 3) summary of key issues in the use of citizen-based analytics in New Zealand (technical issues have been italicised): "the need for social licence and transparency around data use; the need for governance and *data management* structures; the need for *clarity of system architecture; the importance of high-quality data;* and a number of related issues including the data use for service improvement."

CONCLUSION

This chapter applied a novel combination of mixed methods to exploit expert interview data more fully than usual practice. Principles incorporated from the field of expert elicitation enhanced the design of interviews. The interplay between thematic analysis and clustering, facilitated by a crossover mixed design, permitted multi-dimensional insights. While "quali-quantitative" investigations are being increasingly adopted in studies pertaining to "digital sociology," prevention science, and other areas (Henry et al., 2015; Venturini & Rogers, 2019: 536), our chapter tailors a "quanti-qualitative" approach to the analysis of interview data, which is atypical amongst mixed methods studies due to the close integration of quantitative and qualitative findings (Hesse-Biber, 2015:779).

According to the experts we interviewed, data linkage is the new frontier in working with big datasets, in particular, administrative data. Regarding the question of the benefits and challenges of person-centric data linkages, most (but not all) believed that the benefits outweigh the problems, but nearly all had encountered significant challenges in using data linkage to improve the quality of interventions for vulnerable populations. There were very few examples given of actual improvements in health or wellbeing that had resulted from big data analysis. In contrast to an emphasis on volume and complexity, smarter use of existing data, including small, specialised datasets, was believed to add considerable social value. The main challenges were not perceived to relate to the technical or logistic issues involved in linking large datasets, nor to the interpretation of analyses of linked datasets. Rather, the main challenges were identified as the lack of public trust and the risks to privacy, especially for vulnerable children and communities and the data that "represents" them. Overall, these experts felt that we have not yet achieved an appropriate balance between potential benefits and the threats to privacy and human rights, which was our second research question.

Issues of privacy, trust, and sharing (e.g., The Five Safes) have sometimes been addressed via a checklist in a unidimensional way, but these experts mostly emphasised that a more holistic approach is required, in which technical, governance, and rights issues intertwine. This accords with the Northern Ireland experience (Robinson et al., 2018) that public support for data sharing rests on three pillars: trust in organisations, data protection measures, and public benefit. These three pillars were considered interdependent, in that if one crumbles, public support falls away. We may infer that this applies similarly to privacy, trust, and sharing.

It should be noted that for some stakeholders, particularly those in academia, there can be challenges in gaining data custodian approvals required to link data (Homel et al., 2021), although for most other experts we interviewed, this seemed to be less of a concern. A few experts felt that working collaboratively with communities would be a promising approach to overcoming some of the obstacles posed by data linkage. They were of the view that place-based approaches could facilitate respectful dialogue between experts and diverse groups in vulnerable communities, helping to establish trust in the organisations that would have access to data about them, and the ways in which decisions would be made on the basis of these data. Although they did not use the term, what these experts really seemed to have in mind was a process of co-creation (Bason, 2010), which goes beyond "consultation" or "collaboration," in that people with power (the data users, researchers, data experts, service designers) actively share power with data subjects: the power to define the problem in the first place, the power to define "public benefit" in the local context, and the power to agree on the best data ownership and protection strategies and the best kinds of services guided by these data. Although acknowledging data sovereignty was flagged as a potential way forward to address public trust in data linkage, these processes have not, to date, been implemented systematically in Australia, and the infrastructure does not yet exist to do this effectively.

Finally, the experts all agreed that the Australian context for data linkage is at present very fluid. Our sample is relatively small, although results suggest saturation of identifiable themes was achieved (the qualitative equivalent of sample size analysis). Moreover, a sample of 10 experts is considered relatively large when interviewing experts, rather than a more general population. The research design targeted leaders who work closely with government. It is not surprising most of the experts were for, rather than against, the pursuit of data linkage, but the majority agreed that there is an urgent need for a more holistic national framework to underpin data linkage and the use of big data in human services research and service delivery. Furthermore, there needs to be more comprehensive skills development, encompassing not just technical expertise but also the facility to co-design data collection, ownership, analysis, and sharing strategies with the intended beneficiaries of these systems and services.

REFERENCES

Australian Government Department of Education and Training (2018). Better evidence for children's policy. Towards improving children's developmental outcomes: Scoping an approach to build better evidence for policy (Final Report, June 2018). Canberra.

Ayyub, B. M. (2001). Elicitation of expert opinions for uncertainty and risks. Florida, USA: CRC Press.

Bason, C. (2010). Leading public sector innovation: Co-creating for a better society, Bristol: Policy Press.

Biddle, N., Edwards, B., Gray, M., Hiscox, M., McEachern, S., & Sollis, K. (2020). Data trust and data privacy in the COVID-19 period. The ANU Centre for Social Research Methods, Canberra: Australia.

Birchall, C. (2016). Shareveillance: Subjectivity between open and closed data. Big Data & Society, 3(2), 1-12.

Brantingham, P. J. (2017). The logic of data bias and its impact on place-based predictive policing. Ohio State Journal of Criminal Law, 15, 473.

Braun, V., Clarke V., Hayfield N., & Terry, G. (2019). Thematic analysis. In P. Liamputtong (Ed.), Handbook of research methods in health social sciences (pp. 843-860). Singapore: Springer.

Brod, M., Tesler, L., & Christensen, T. (2009). Qualitative research and content validity: Developing best practices based on science and experience. Quality of Life Research, 18(9), 1263-1278.

Brown, A. P., Ferrante, A. M., Randall, S. M., Boyd, J. H., & Semmens, J. B. (2017). Ensuring privacy when integrating patient-based datasets: New methods and developments in record linkage. Frontiers in Public Health, 5, 1-6.

Brownell, M. D., & Jutte, D. P. (2013). Administrative data linkage as a tool for child maltreatment research. Child Abuse & Neglect, 37 (2-3), 120-124.

Caley, M. J., O'Leary, R. A., Fisher, R., Low Choy, S., Johnson, S., & Mengersen, K. (2014). What is an expert? A systems perspective on expertise. Ecology and Evolution, 4 (3), 231-242.

Carney, T. (2019). Robo-debt illegality: The seven veils of failed guarantees of the rule of law? Alternative Law Journal, 44 (1), 4-10.

Ceri, S. (2018). On the role of statistics in the era of big data: A computer science perspective. Statistics & Probability Letters, 136, 68-72.

Choi, S. S., Cha, S. H., & Tappert, C. C. (2010). A survey of binary similarity and distance measures. Journal of Systemics, Cybernetics and Informatics, 8(1), 43-48.

Cox, D. R., Kartsonaki, C., & Keogh, R. H. (2018). Big data: Some statistical issues. Statistics & Probability Letters, 136, 111-115.

Cuccaro-Alamin, S., Foust, R., Vaithianathan, R., & Putnam-Hornstein, E. (2017). Risk assessment and decision making in child protective services: Predictive risk modeling in context. Children and Youth Services Review, 79, 291-298.

Dankar, F. K., El Emam, K., Neisa, A., & Roffey, T. (2012). Estimating the re-identification risk of clinical datasets. BMC Medical Informatics and Decision Making, 12 (1), 1-15.

Department of Education, Skills and Employment (2021). Complete privacy policy. Canberra: Australian Government.

Department of the Prime Minister and Cabinet (2019). Data sharing and release legislative reforms. Discussion Paper. Canberra: Commonwealth of Australia.

Department of Science, Information Technology and Innovation (2016). Queensland Government open data policy statement. Canberra: Queensland Government. https://www.data.qld.gov.au/_resources/documents/qld-data-policy-statement.pdf

Desai, T., Ritchie, F., & Welpton, R. (2016). Five Safes: Designing data access for research. Economics Working Paper Series. Bristol: University of the West of England.

de Souza, Y. (2014). Not just data: Privacy in the digital age. Feliciter, 60(5), 17-22.

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. Journal of Travel & Tourism Marketing, 12(1), 1-22.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. New York: St. Martin's Press.

European Commission (2017). Enter the data economy. EU policies for a thriving data ecosystem. EPSC Strategic Notes, 21, 1-16.

Falk, A. (2019). Data sharing and release. Legislative Reforms Discussion Paper, Submission to Prime Minister and Cabinet. Press Release. Canberra. www.oaic.gov.au

Fisher, C. B., & Layman, D. M. (2018.) Genomics, big data, and broad consent: A new ethics frontier for prevention science. Prevention Science, 19 (7), 871-879.

Friedman, J., Hastie, T., & Tibshirani, R. (2009). The elements of statistical learning (2nd ed.). New York, Springer.

Galloway, K. (2017). Big data: A case study of disruption and government power. Alternative Law Journal, 42 (2), 89-95.

Gluckman, P. (2017). Using evidence to inform social policy: The role of citizen-based analytics. Auckland: Office of the Prime Minister's Chief Science Advisor. http://www.pmcsa.org.nz/wpcontent/uploads/17-06-19-Citizen-based-analytics.pdf

Green, M. J., Harris, F., Laurens, K. R., Kariuki, M., Tzoumakis, S., Dean, K., . . . Carr, V. J. (2018). Cohort profile: The New South Wales child development study (NSW-CDS) – Wave 2 (child age 13 years). International Journal of Epidemiology, 47(5), 1396-1397k.

Henry, D., Dymnicki, A. B., Mohatt, N., Allen, J., & Kelly, J. G. (2015). Clustering methods with qualitative data: A mixed-methods approach for prevention research with small samples. Prevention Science, 16, 1007-1016.

Hesse-Biber, S. (2015). Mixed methods research: The "thing-ness" problem. Qualitative Health Research, 25(6), 775-788.

Holman, C. D. A. J., Bass, J. A., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J., . . . de Klerk, N. H. (2008). A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. Australian Health Review, 32(4), 766-777.

Homel, J., Homel, R., McGee, T. R., Zardo, P., Branch, S., Freiberg, K., Manning, M., & Wong, G. (2021). Evaluation of a place-based collective impact initiative through cross-sectoral data linkage. Australian Journal of Social Issues, 56(2), 301-318.

Hopf, Y. M., Bond, C., Francis, J., Haughney, J., & Helms, P. J. (2014). Views of healthcare professionals to linkage of routinely collected healthcare data: A systematic literature review. Journal of the American Medical Informatics Association, 21(e1): e6-e10.

Kaminski, M. E. (2019). The right to explanation, explained. Berkeley Technology Law Journal, 34(1), 189-218.

Korda, R., Di Law, H., Armstrong, B., D'Este, C., Smurthwaite, K., Trevenar, S., Lucas, R., Miller, A., & Kirk, M. The PFAS Health Study: Data linkage study research protocol. https://rsph.anu.edu.au/files/

Knol, A. B., Slottje, P., van der Sluijs, J. P., & Lebret, E. (2010). The use of expert elicitation in environmental health impact assessment: A seven step procedure. Environmental Health, 9(1), 1-16.

Križan, T., Brakus, M., & Vukelić, D. (2015, May). In-situ anonymization of big data. In 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 292-298.

Kukutai, K., & Walter M. (2015). Indigenising statistics: Meeting in the recognition space. Statistical Journal of the IAOS, 31(2), 317-326.

Kukutai, T., & Taylor, J. (2016). Data sovereignty for indigenous peoples: Current practice and future needs. In T. Kukutai & J. Taylor (Eds.), Indigenous data sovereignty (pp.1-22). Canberra: ANU Press.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. Business Horizons, 60(3), 293-303.

Leonard, D., Buettner, P., Thompson, F., Makrides, M., & McDermott, R. (2018). Linking 'data silos' to investigate anaemia among Aboriginal and Torres Strait Islander mothers and children in Far North Queensland. Australian and New Zealand Journal of Public Health, 42 (5), 456-462.

Low-Choy, S., Almeida, F., & Rose, J. (2021). Combining study findings by using multiple literature review techniques and meta-analysis: A mixed-method approach. In E. Manu & J. Akotia (Eds.), Secondary research methods in the built environment (Chapter 15). Oxford: Routledge.

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. New Media & Society, 19(5), 780-794.

McNamara, B., Gubhaju, L., Jorm, L., Preen, D., Jones, J., Joshy, G., & Eades, S. (2018). Exploring factors impacting early childhood health among Aboriginal and Torres Strait Islander families and communities: Protocol for a population-based cohort study using data linkage (the 'Defying the Odds' study). BMJ Open, 8(3), 1-12.

Munro, E. (2019). Predictive analytics in child protection. CHESS Working Paper No. 2019-03. Durham University.

NSW Government (2021). NSW Data Analytics Centre. https://data.nsw.gov.au/nsw-data-analytics-centre

Office of the National Data Commissioner (2020). Data Availability and Transparency Bill 2020: Exposure draft consultation paper. Canberra: Department of the Prime Minister and Cabinet.

Onwuegbuzie, A. J., & Hitchcock, J. H. (2015). Advanced mixed analysis approaches. In S. Hesse-Biber & N. Johnson (Eds.), The Oxford handbook of multimethod and mixed methods research inquiry (Chapter 16). Oxford University Press.

Piotrowska, P. J., Whitten, T., Tzoumakis, S., Laurens, K. R., Katz, I., Carr, V.J., Harris, F., & Green M. J. (2020). Transitions between socio-emotional and cognitive vulnerability profiles from early to middle childhood: A population study using multi-agency administrative records. European Child and Adolescent Psychiatry, 29, 1659-1670.

Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. International Journal of Nursing Studies, 47(11), 1451-1458.

Productivity Commission (2017). Data availability and use (Report No. 82). Canberra.

Quarteroni, A. (2018). The role of statistics in the era of big data: A computational scientist's perspective. Statistics & Probability Letters, 136, 63-67.

Queensland Health (2020). Queensland data linkage framework. State of Queensland, Brisbane. https://www.health.qld.gov.au/__data/assets/pdf_file/0030/150798/qlddatalinkframework.pdf

QSR (2017). About cluster analysis, in NVivo 11 for Windows. QSR International. https://helpnv11.qsrinternational.com/desktop/concepts/about_cluster_analysis.htm

QSR (2018). NVivo (Version 12 Pro). QSR International Pty Ltd. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Robinson, G., Dolk, H., Dowds, L., Given, J., Kane, F., & Nelson, E. (2018). Public attitudes to data sharing in Northern Ireland: Findings from the 2015 Northern Ireland Life and Times survey. Ulster University. https://www.ark.ac.uk/pdfs/Researchreports/Data_Sharing_Report.pdf

Rose, J., Low-Choy, S., Singh, P., & Vasco, D. (2020). NAPLAN discourses: A systematic review after the first decade. Discourse: Studies in the Cultural Politics of Education, 41(6), 871-886.

Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Chouldechova, A., & Diaz, F. (Program Chairs) Proceedings of the Conference on Fairness, Accountability, and Transparency, Jan 29-31, 2019, Atlanta, 59-68. NY: Association for Computer Machinery.

Sexton, A., Shepherd, E., Duke-Williams, O., & Eveleigh, A. (2017). A balance of trust in the use of government administrative data. Archival Science, 17, 305-330.

Tam, S. M., & Kim, J. K. (2018). Big data ethics and selection-bias: An official statistician's perspective. Statistical Journal of the IAOS, 34(4), 577-588.

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley Publishing Company.

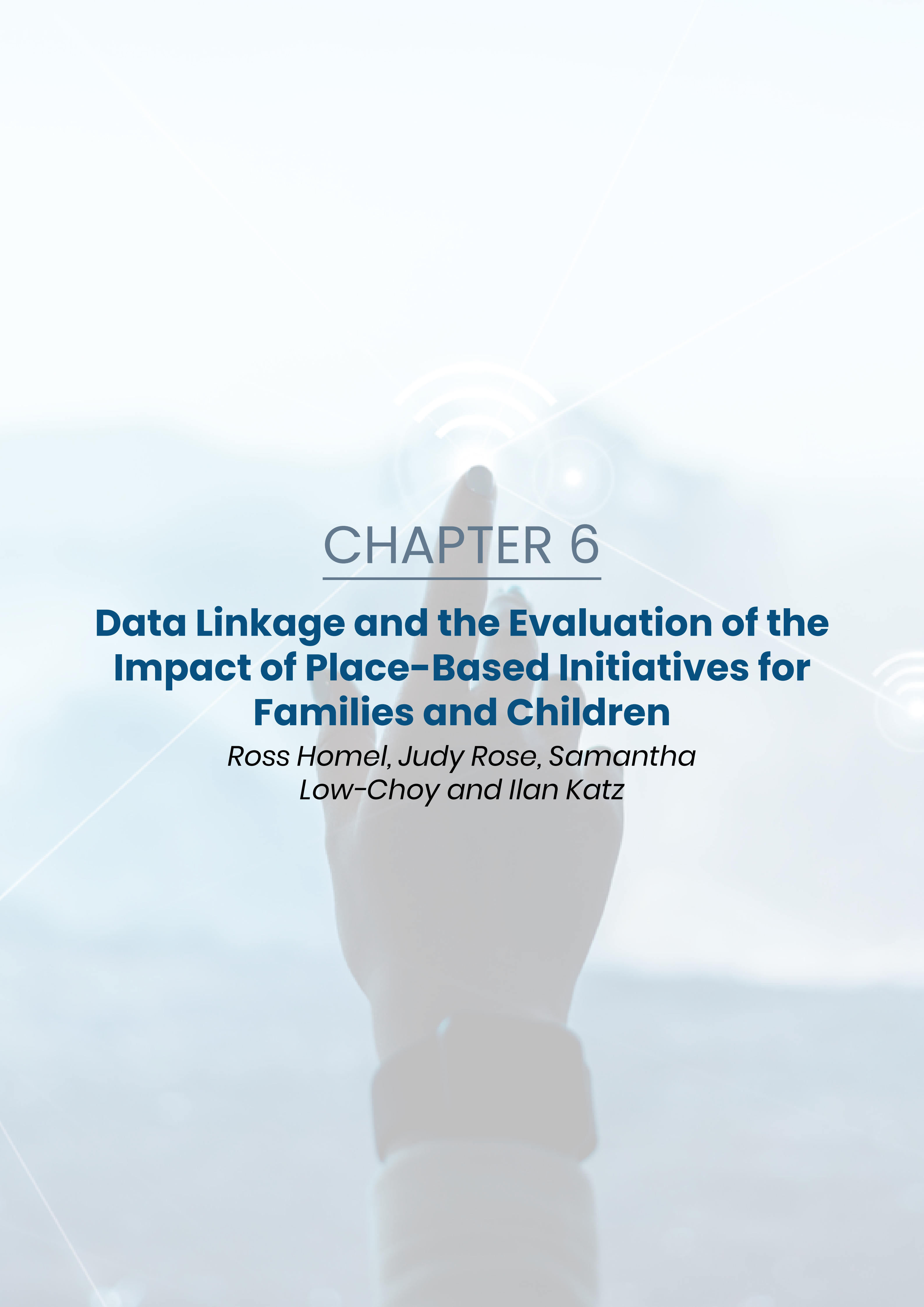
Vaithianathan, R., Maloney, T., Putnam-Hornstein, E., Jiang, N. (2013). Children in the public benefit system at risk of maltreatment: Identification via predictive modelling. American Journal of Preventive Medicine, 45(3), 354-359.

Venturini, T., & Rogers, R. (2019). “API-based research” or How can digital sociology and journalism studies learn from the Facebook and Cambridge Analytica data breach. Digital Journalism, 7(4), 532-540.

Walter, M. (2018). The voice of indigenous data: Beyond the markers of disadvantage. Griffith Review, 60, 256-263.

Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. International Journal of Social Research Methodology, 22(3), 233-243.

World Economic Forum (2019). Data collaboration for the common good: Enabling trust and innovation through public-private partnerships. Geneva: Switzerland: McKinsey & Company.



CHAPTER 6

Data Linkage and the Evaluation of the Impact of Place-Based Initiatives for Families and Children

Ross Homel, Judy Rose, Samantha Low-Choy and Ilan Katz

Data Linkage and the Evaluation of the Impact of Place-Based Initiatives for Families and Children[††]

INTRODUCTION

> *Organizations are now linking and connecting diverse datasets at an accelerating pace to create value—and this is one of the primary factors shaping today's global economy. … As it relates to the global development and humanitarian sectors … [a] daunting set of challenges lies in bringing together small amounts of non-standardized "public-to-public" data from dozens of NGOs and organizations to create a common picture of needs and response. (World Economic Forum, 2019: 8)*

Across the Western world there has been a social shift from centralised, top-down, state-controlled policies and services toward participative, bottom-up, outcomes-oriented and participant-managed approaches that "give voice" to strongly felt needs and address the priorities of communities and vulnerable groups (Wallace, 2019). One manifestation of this trend is the proliferation of place-based initiatives (PBIs) which focus on improving the social conditions of communities through the combined activities of government, local, and non-government organisations (Homel et al., 2021). Place-based (or area-based) initiatives focusing on children's wellbeing became prominent in the 1990s with the implementation of the Sure Start program in the UK (Eisenstadt, 2011). The approach was adopted earlier in Australia (Muir et al., 2009) with the development of the Communities for Children program in 2004, and a range of similar initiatives have followed.

This approach has been met with enthusiasm by communities and by many parts of the community services sector (Paulson & Bailey, 2018), but with some ambivalence by governments. Arising perhaps from lingering doubts about the value of place-based programs that may touch only a portion of the population, governments are coming to the realisation that policy in this area is poorly developed and that there is a chronic lack of infrastructure to support place-based initiatives. Given the policy challenges related to implementing PBIs and taking them to scale, evaluations of these programs are crucial for developing a robust evidence base for their effectiveness and an understanding of the contexts in which improvements in the wellbeing of disadvantaged children and families are most likely. However, there are significant challenges to conducting rigorous evaluations of PBIs, often related to the availability and quality of data, and so evaluators are increasingly turning to linked administrative data.

In this chapter, we explore some of the changes required in the forms of data that are available, and the restrictions placed upon them, in the context of the urgent need to develop replicable methodologies to make possible the rigorous evaluation of place-based initiatives and to better assess implications for evaluation of social policy. The rationale for this is that Australia and many other countries spend hundreds of millions of dollars each year on services for children and families in socially disadvantaged areas. Many programs and initiatives are evaluated, but almost nothing is known about the combined impact on child wellbeing of community-based and school-based programs. In typical socially deprived communities in most rich countries there are likely to be a large number of programs, projects and initiatives for children and families, based in community centers, schools, sports facilities and so on. Some of these initiatives may be effective for some or most of the children that access them, while others may be less effective in the context of particular communities for particular groups of children. But how do these initiatives interact with each other? What is their combined impact on the life course trajectories of children and families living in those communities?

We draw in our reflections on interviews with the 12 experts who were introduced and described in Chapters 4 and 5. These experts provided important insights into the progress made and the many remaining challenges of linking big datasets to improve services to people, including those from vulnerable groups such as children in socioeconomically disadvantaged communities.

ADVANTAGES OF PLACE-BASED INITIATIVES

PBIs offer a number of advantages over traditional preventive or early intervention approaches to improve the wellbeing of vulnerable children. Firstly, these initiatives focus on areas of disadvantage

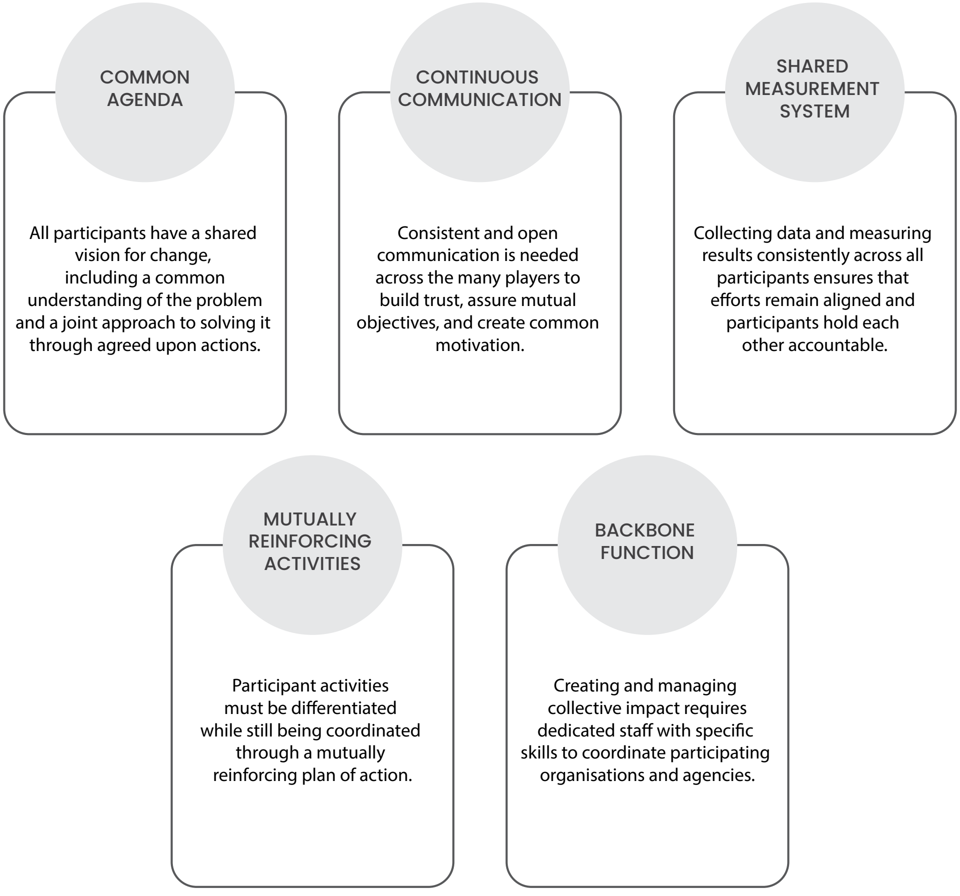
---

[††] An expanded version of this chapter, focused on child maltreatment, will be published as Katz et al. (2021, forthcoming). There is therefore considerable overlap between the text of these two chapters.

where children suffer a range of challenges to their wellbeing within their families and in the community. Multiple risk factors can therefore be targeted within the interventions for a range of children in the community. Traditional approaches to early intervention tend to be fragmented and “siloed,” focusing on specific risk factors (e.g., inadequate housing, disadvantaged schooling, poor parenting), particular mental and physical conditions (e.g., autism spectrum disorder, obesity, suicide), or particular outcomes (child maltreatment, educational attainment, crime). Place-based initiatives aim to “join up” service provision so that children and families in disadvantaged communities are provided easy access to preventive services and are not confined to one or another pathway into support, depending on diagnosis or specific circumstances. Another advantage of this approach is that because the initiative is usually community wide, it does not stigmatise individual children and families who have specific needs, although communities themselves can still be stigmatised. In addition, these initiatives not only provide better services in the community but are designed to mobilise and empower community members to take collective responsibility for improving the wellbeing of local children. However, there remains some ambiguity in the notion of a place-based initiative. This relates to the question of whether the initiative involves a range of services which happen to serve a particular geographic community or whether they primarily take a community development approach, which focuses on improving conditions for the whole community where targeted services were not the primary delivery mode.

Many of the more recent place-based initiatives for children address this conundrum by utilising the “collective impact” approach (Figure 6.1). The core of collective impact involves the range of stakeholders in a community, including local policymakers, service providers, community leaders and community members themselves coming together to address a particular problem or concern for the community. This results in a common agenda, where stakeholders agree on the priority issues to be resolved and the most effective approach for addressing these issues. However, this is typically not sufficient to achieve success. Collective action requires this common agenda to be underpinned by a clear plan and a measurement system so that it is apparent to all stakeholders how things are changing in the community. In addition, a key component of collective impact is the “backbone function,” which usually takes the form of an agency or a group of people who are dedicated to the coordination of activities and interventions in the community and the provision of support in the form of technical and human infrastructure.

Figure 6.1. Collective impact principles



All participants have a shared vision for change, including a common understanding of the problem and a joint approach to solving it through agreed upon actions.

Consistent and open communication is needed across the many players to build trust, assure mutual objectives, and create common motivation.

Collecting data and measuring results consistently across all participants ensures that efforts remain aligned and participants hold each other accountable.

Participant activities must be differentiated while still being coordinated through a mutually reinforcing plan of action.

Creating and managing collective impact requires dedicated staff with specific skills to coordinate participating organisations and agencies.

Source: Preskill et al. (2014: 4)

PBIs often follow a basic model: funding is allocated to a lead agency which uses part of the funding to coordinate the initiative and distributes funding to other organisations to provide a range of services and supports to the community. Sometimes the lead agency provides the backbone function but this can also be provided by another agency.

Although PBIs offer the prospect of significant improvements on traditional interventions for vulnerable children and families, they face some significant challenges. Firstly, the concept of “place” in PBIs can differ significantly between programs or even within programs. Initiatives can focus on very small areas, even down to city blocks, but generally focus on suburbs or bigger areas such as municipal boundaries like towns or local government areas. Rural areas tend to be much larger geographically but smaller in population than urban areas, which may create challenges for some children and families to access the range of services and supports offered by the initiative. Moreover funding levels may differ between areas.

The imperative for cross-portfolio linked data to be available to communities and community-based organisations, one of the central requirements of a collective impact approach (Kania & Kramer, 2011), poses particular challenges. As a 2017 report by the Australian Productivity Commission put it:

> *Lack of trust by both data custodians and users in existing data access processes and protections and numerous hurdles to sharing and releasing data are choking the use and value of Australia’s data. In fact, improving trust community-wide is a key objective.* (Productivity Commission, 2017: 2)

EVALUATIONS OF PLACE-BASED INITIATIVES

Although there have been a number of evaluations of place-based initiatives, it remains extremely challenging to conduct robust evaluations in this field. Traditional experimental and quasi-experimental approaches, which are the most rigorous methods for evaluating programs or services, are designed primarily for evaluations of specific programs or interventions. In the most rigorous evaluations, every participant in the intervention group receives a similar experience/dosage of the intervention, and their outcomes in specific areas or domains are compared with those of similar people who do not receive the intervention (controls). The “gold standard” methodology for these designs is a randomised controlled trial (RCT) in which participants are randomly assigned to the intervention and the control groups. This design is intended to eliminate bias and is the best way to ensure that the outcomes measured are attributable to the intervention rather than to other factors which may influence these outcomes of interest. Quasi-experimental designs are less rigorous, but similarly compare participants in the intervention with other people who do not participate.

These methods are challenging to apply to PBIs for a number of reasons that are discussed below. Nevertheless, it is important to note that in the United States, many of the challenges we discuss have been overcome, and complex PBIs have been successfully evaluated through both RCTs and population-level quasi-experimental designs, using communities (such as counties) as the prime unit of analysis, supported by evaluations at the individual level. The prime example is the Communities That Care community mobilisation and evidence-based intervention system which has been evaluated through a 24-community RCT across seven states (the Community Youth Development Study), as well as through a 5-year effectiveness trial across most of the counties in Pennsylvania, using a matched-counties design (Fagan et al., 2019). Both evaluations showed large, cost-effective, and sustained impacts on youth behavioural health problems at the community level, including crime, substance abuse, and school dropout. Because we do not yet have the same tradition in Australia of a commitment to quantitative methods and the use of experimental designs in the community services sector, we offer the following discussion and reflections on the views of experts as a contribution to the development of an evaluation reform agenda in this country.

A particular challenge for evaluations of PBIs is to find a suitable comparison or control group with which to benchmark outcomes for the intervention area. A fundamental problem is that, unlike interventions such as parenting programs or early learning interventions, PBIs, like Communities That Care or Communities for Children (Homel et al., 2015) are tailored to each community and may be very different in different locations. Indeed the “dosage” (funding and resourcing) of the intervention may differ considerably between intervention sites.

Another challenging issue is that communities are generally specifically chosen for the initiatives because of a range of factors, including economic, demographic and geographic characteristics. Although it may, in principle, be possible to randomly allocate place-based initiatives to communities, this is generally not feasible practically or politically. Communities are selected because they are disadvantaged and lack coordinated services to meet the needs of vulnerable children. If the initiative is based on the Collective Impact Model described above, the community must have the basic infrastructure in place to coordinate responses, share information and provide backbone support. Similarly, residents are not randomly allocated to services. These can vary from intensive therapeutic programs for a small number of children through to projects such as fun days in open spaces, volunteering activities for parents, breakfast and homework clubs for children, and so on. Each program within the community will target a particular group of children and families, and each child within a community may receive a different range of interventions, and different dosages of each intervention. This may be because of different needs being served by different combinations of interventions, but there could be other reasons for children receiving different services, such as accessibility or appropriateness of services for particular sub-populations.

Perhaps the key conceptual issue for the evaluation of PBIs, alluded to earlier, is the unit of analysis: should outcomes be measured at the child or community level? The logic behind collective impact approaches is that all children in the community should benefit, a logic adopted by Communities That Care, which is itself a type of collective impact initiative. Therefore, outcomes should be measured for all children in the initiative site or at least for a representative sample, and these should be compared to children in similar areas which do not receive the intervention. However, comparing average outcomes in intervention areas with those of children in other areas can obscure as much as it reveals. For instance, this approach would not allow any analysis of which particular children in the community received what services, and over what period of time. It is theoretically possible, for example, that Area A, which does receive the program, improves significantly compared to Area B which does not, but that the improvement may be entirely driven by children in the community who had no contact with any of the funded services involved in the initiative. It may also be that services in Area A are poorly targeted, and those children who most need the intervention are missing out, or, conversely, that there is an overlap of services for some issues and lack of services for others. Thus, it is important for evaluations not just to measure the outcomes of interest for the intervention group compared to the comparison group, but to track the trajectories of service use of children and families and link their access to different services over time to their outcomes (Homel et al., 2021).

A different kind of problem for evaluations arises because many low socioeconomic status (SES) communities are characterised by high levels of mobility. Community residents can move in and out of the community over time. Mobility is a particular issue for many disadvantaged communities (Edwards et al., 2014). Thus, some children may be living in the community only for short periods and they will have limited exposure to the initiative. Further, children may enter and leave the community at different stages of their lives. Similarly, children in the community may receive services or supports outside of the community. It is therefore very difficult to know who has received what intervention over what period.

Because of these challenges, it is very difficult and expensive to track service use over time for children and families in the community using traditional research methods (surveys, interviews etc.), and to track children and families in communities which have not received the intervention. Some projects have used this methodology, including Communities That Care, that routinely uses a comprehensive youth survey administered through schools, the evaluations of Sure Start in the UK (Melhuish et al., 2010) and Communities for Children in Australia (Edwards et al., 2014). However, each of these evaluations had limitations—they compared outcomes for a sample of children in the intervention communities with children in comparison communities which did not have the program. These evaluations were not able to track participants' access to services and other interventions, and therefore could not attribute the outcomes to the specific interventions which the children had been engaged with. In some ways, this is consistent with the collective impact philosophy, in that it hypothesises that all children in the communities with collective impact should benefit. However, this still leaves a significant gap in understanding what specific components of the interventions appear to be more or less effective in improving outcomes for children. Indeed, in all cases, there was a great deal of variation within the communities as well as between them, indicating that there may be specific combinations of factors which may positively influence outcomes, rather than a particular program model.

These evaluations highlight the fact that even when community members are surveyed over time and compared with children and families in similar communities, interpreting the findings can be very difficult because of the complexity of communities and interventions. In particular, collective impact involves not only improved services but also communities acting together (joined-up services) and better use of data. If improvements are found in the outcomes of children in that community, it is not easy to say which component of the initiative has "caused" this—better services, more joined-up services, better planning and collaborative service delivery, a combination of these factors, or simply that these communities have received additional funding over and above the normal service allocations.

POTENTIAL ROLE OF DATA LINKAGE

Advances in technology and data analytic techniques have made it possible to link many different datasets with large numbers of variables and data subjects. Collection, linking and analysis of administrative data from multiple service providers has therefore come to be seen as a potential solution to some of the many problems in the implementation and evaluation of place-based initiatives, and indeed, many other types of interventions.

All agencies that provide support to children and their families collect administrative data about occasions of service and the type of service provided, and some agencies also routinely assess children on a range of outcomes, including physical, social and emotional development, health, and cognitive development/educational attainment. It is therefore theoretically possible for the service trajectories of every child in the community (rather than only a sample) to be tracked over time. Outcomes can be measured through administrative data, in particular, by linking service data to assessments such as NAPLAN scores, school attendance, health assessments and other administrative outcomes such as criminal justice involvement, employment, Emergency Department presentations and homelessness.

These data could then be compared to similar children (or children in similar communities) so that the impact of the intervention can be compared. This type of evaluation therefore provides very significant advantages over primary data collection through surveys, interviews or other traditional evaluation methods. This method also reduces the burden on participants because families or young people do not have to complete questionnaires. Data linkage is much less expensive than surveying participants over several waves of data collection. Linked data can track the trajectories of children and parents over long periods, and therefore the impact of the interventions can be tracked longitudinally over many years.

Nevertheless, the logistical, ethical and conceptual challenges to this approach are formidable, as discussed in Chapters 4 and 5. Some of the issues are:

- Agencies collect data in different ways using different platforms.
- The consent obtained from data subjects differs and often precludes the data being used for purposes other than that for which they were collected.
- Confidentiality can be a significant issue in small communities where data subjects could easily be identified.
- Data can be of poor quality, with key variables missing in non-random ways. For example, some fields such as "language spoken at home" may be less accurately completed than other fields.
- Interpretation of findings can be very difficult and often has to rely on detailed knowledge of the community context.
- Finding comparison communities or other units of analysis (schools, hospitals, day care centres) that are sufficiently similar to the intervention communities or units can be difficult.

One of the main challenges for the use of linked data in evaluations is that because the data are primarily used for administrative purposes, they measure events such as doctors' appointments, police contacts, and school attendance, and not the "softer" or less measurable outcomes such as parenting style, family relationships or self-esteem. It is, of course, possible to combine administrative data with surveys, and that is an approach that is increasingly being used, as described in Homel et al. (2021).

WHAT EXPERTS SAY

General challenges

Experts interviewed for this study suggest that making data more linkable would improve inter-operability between agencies, especially those that work with vulnerable groups. One expert outlined the issue of interoperability:

> *In establishing a national operating environment that's interoperable moving beyond the linked silos that we have and still lets everyone have their autonomy and bring their own little neck of the woods together, but also makes it interoperable so that people can actually make good use of it by being able to bring different components together quickly and easily with the minimal of cost and time spent. (Expert #1)*

Another expert discussed the possibility of data linkage at a national, state and local level, noting that it is "easier" to use a place-based approach:

> *You could use data linkage at a whole of state level or a whole of Australia level, or you could just do it for a defined population of children, say Indigenous children, and they could be everywhere. But it's easier if you use a place-based approach, because you've got these agencies already meeting in a locality, often in a school … There's a lot of convenience, a lot of benefit in working in a locality, in a place-based, what we call in Australia a place-based way or community-based approach. It's not absolutely essential; it could be a more dispersed population that is the focus of these sorts of linkages. I've talked a lot about the very broad field of community prevention. I just want to get back to the need to build the infrastructure to do this, the scientific infrastructure, of which a critical part is the capacity to collect high-quality data on what's happening with these kids in terms of the funded programs and services in schools and elsewhere. (Expert #7)*

Another couple of experts also noted that in disadvantaged communities there is a problem of unknown dimensions, with the duplication of services for some children, while others "fall through cracks" despite high need. This problem is outlined in detail by one expert, indicating that it is difficult to evaluate outcomes of place-based initiatives when it is not clear who is being serviced and for how long:

> *So, I want to know how many people have got service agreements through funding arrangements that say they are funded for that postcode to support children and families' wellbeing. Is there then duplication, or are there gaps? Is there no one providing that support required? And I'd also want to know: actually, could we see how much money is spent on each family? Or we've got too many people going to some families and they're not the right families to support? (Expert #4)*

Another expert reflects on this problem, but brings in a broader human rights perspective to communities which may be "disenfranchised" or "alienated" in the process of receiving such services.

> *Are our kids being deprived in the system because they're falling between the cracks or are they becoming the subject of receiving too much money, because they're getting it from multiple sources, and really it should be better allocated and so they're concerned that they might actually experience this slicing approach that takes away things from them? … You might be doing it now or in an ad hoc fashion, but there has been no mention of an upfront risk assessment either of human rights, or, in my case, privacy rights at all, or how this is going to actually disenfranchise or alienate the communities that you're most seeking by these definitions of disadvantage want to include. And I would say that the Indigenous communities inside Australia would be the ones who are most likely going to be put off by this type of certain paternalistic assumptions. (Expert #11)*

Picking up on this theme, another expert noted that care needs to be taken with place-based evaluations not to identify a community or individuals living within such a community.

> *I think there, you do have the extra problem of confidentiality because most place-based initiatives are like one suburb or whatever. If you think, well, how many under-fives are in [disadvantaged suburb] and of those under-fives, how many have a disability? … It's quite easy to identify individuals. So, you've got to be careful of that sort of thing when you're dealing with place-based. But again, whether you're using linked data or admin data versus surveys, it's not that dissimilar. … Obviously, as a researcher, I think overall, it's for the public good, but that doesn't mean that there aren't issues around things like privacy and consent and other questions. (Expert #8)*

Some experts noted that Australia needs to build better modelling capacity in the analysis of big data collected on vulnerable communities. This entails evaluating existing data collected on place-based initiatives to better guide future needs, as explained by another expert:

> *Look, on the data, I think I've been around long enough to have seen and run place-based initiatives. Most of them [are] around in the same places, so how you ever really decided which place-based initiative in the same place had what impact is a bit of a tough ask. Somewhere like [disadvantaged suburb in metropolitan city] has had 15, 20, 25 years of place-based initiatives. But the thing that puzzles me is people don't ask about how can you look at the data that already exists to understand where something might be happening that's a bit different in a particular location without doing a new place-based initiative. Then, heavens forbid, go and ask them and say it seems to be working here differently to somewhere else. Why is that so? Try and find somewhere where they've just worked out for themselves what to do differently. The data will be in there. You've got 20 years plus of de-identified data, so you'd be able to find what kind of responses have been done for young carers, for kids coming out of juvenile justice, kids from out-of-home care, teenage parents. … But, of course, data work isn't as sexy as putting two or three million dollars into a particular location. (Expert #10)*

Another set of data design issues for evaluating place-based initiatives was noted by Expert #2:

> *With any place-based evaluation, you've still got the problem of what's the comparator group. So you've got to get data, ideally, I suppose, historically as well, to see whether the curve changes in place-based initiatives. … If the data on the outcomes is already there being collected, then I suppose it helps a place-based initiative. Who then has to try and find a comparator place and get data from the comparator place? I think that's always a challenge logistically and financially. If the relevant outcome information was there, then that would be a relatively trivial exercise if you've got a large link data, whole of population data platform. That's also got a strong historical component, so you can pick up historical trend. Then hopefully, what you see is that, in the intervention communities, you see a change in the rate or the prevalence of something that you're interested in. Of course, you don't see it in the other communities, and that becomes a strong research design. I think the potential for that is certainly there, but I think it would be a relatively cost-effective way to utilise those things, assuming that you've got the outcomes that you're interested in. That's always the challenge: do you have the outcome that the intervention is meant to change in the community? (Expert #2)*

In summary, interviews with experts suggest that challenges remain in terms of data quality and availability, organisational capacity, technical infrastructure and legal/ethical considerations among organisations that collect data on humans and the various services they receive, including at a place-based level. However, organisations that have well-defined governance procedures around linked datasets are making some progress, as well as those that use a place-based approach to build trust around the use of data at an individual and community level. Running through many of the interviews was the fundamental need to find a balance between community trust and technical innovation. This was confirmed by the more detailed analysis conducted in Chapter 5 around the way that trust, privacy and data sharing were intertwined in the experts' minds. The balance issue also permeates the findings of Chapter 4, where experts observed that some reforms are being made in light of lessons learned from high-profile data linkages such as Robodebt. These lessons include that misuse is related to measures designed to develop trust (communication, transparency, accountability and legality; Chapter 4, Lesson 2), and in particular that accountability must go hand in hand with technical efforts (Chapter 4, Lesson 4).

In this chapter, we show how a range of technical and ethical barriers to data linkage and analytics persist even in collaborative place-based initiatives where the focus is on data at the community level, not the individual level. The key challenge from the perspective of the experts was the deep mistrust of data analysts expressed by both members of the public and by providers of data beyond the community. These sceptical views have been influenced by media reports of the Robodebt scandal and other incidents in Australia where data has been misused by government, as well as incidents where personal records have been hacked. The place-based approach has a key role to play in supporting residents in communities to trust the evaluators and to work with them to ensure that data are used appropriately without stigmatising or damaging individuals or the communities as a whole.

Trust issues

While trust, privacy and ethics around sharing sensitive data are important principles, they are difficult to apply in practice. A number of challenges have to be overcome for data to be used successfully and ethically for evaluations.

As one of our experts explained:

> *Not having access to the data, to getting all the approvals in place and making it happen, I think that really is the big issue. There's a rich variety of data out there. We've certainly got the technical capability to bring it all together and we can bring it all together securely, but the real problem is that … we have to go through the rigmarole of a number of ethics committees, a number of data custodians that you need to go through, and then their willingness to be able to do it. (Expert #1)*

This quote illustrates other areas of trust that must be negotiated for successful evaluations. Those agencies which hold datasets have to trust the evaluators that they will use the data appropriately and responsibly, both in terms of how the data is stored, but also how it is used and reported. In particular, it is crucial that individuals are not able to be identified in the analysis. Secondly, the evaluators themselves must trust the agency to provide the data and also trust that the data are fit for purpose for the evaluation. Data must therefore be as reliable and complete as possible and be appropriate, given the research questions.

In particular, place-based evaluations may involve relatively small numbers of participants and therefore significant amounts of missing data can be very problematic for evaluations, particularly if there are systematic patterns in the missing data. For example, if children who are most vulnerable or have most problems at school tend to miss school tests or exams, this can give a misleading impression of their outcomes. Similarly, as mentioned earlier, families who are functioning optimally tend to move out of the area, which can affect the analysis unless the children are able to be tracked in the dataset to other locations. A further important issue with this type of data analysis is that although it gives the impression of objectivity, there are often biases built into the data themselves, and these can provide misleading or even stigmatising findings (Krakouer et al., 2021). This is because the decisions about what data to collect, how to collect it, and how to record it are made by people who may be unaware of their own biases and assumptions.

CONCLUSIONS AND POLICY IMPLICATIONS

There is a near consensus that using data linkage to evaluate place-based initiatives for children is a worthwhile idea. However, to date, this has not been done in Australia, and there are very few international examples, because there continue to be a range of technical, ethical and governance issues for these evaluations to be conducted effectively and efficiently. Experts suggest that efforts to date have been piecemeal and fragmentary, and that progress will require national leadership. Using linked data requires both technical expertise and policy/practice understanding, and this expertise is still being developed.

We propose that place-based approaches could be adopted to overcome some of these barriers, by allowing efforts around data collection and linking to be respectful of ethical and individual dignity, by working with and not on people. Achievement of this goal may require "smart" thinking via a targeted strategic data linkage approach. Another key element of this approach is to engage with community leaders and community members throughout the intervention and the evaluation and to work with them to co-design data collection, analysis, ownership, and dissemination. Engagement with communities is crucial for the development of trust in the use of linked datasets for the benefit of the community. Similarly, as Homel et al. (2021) point out, it is equally important to engage at an early stage with data custodians and other stakeholders to ensure that datasets are available, can be accessed ethically and are complete and sufficiently reliable to use for evaluation purposes.

In conclusion, addressing the technical feasibility and legal and ethical appropriateness of linking multiple datasets from diverse sources, including state, Catholic and independent primary schools and education department records; state-funded intensive family support services; and Commonwealth-funded services would truly break new ground in Australia. Unlike most existing datasets that link administrative data about children,[‡‡] what this chapter explores is the feasibility of linking data longitudinally on patterns of participation in programs or services designed to improve child wellbeing with a novel range of child outcomes, such as teacher-reported behavior or social-emotional wellbeing as reported by the children, giving them a direct voice in the data generation process (Day et al., 2019).

REFERENCES

Day, J., Freiberg, K., Hayes, A., & Homel, R. (2019). Towards scalable, integrative assessment of children's self-regulatory capabilities: New applications of digital technology. Clinical Child and Family Psychology Review, 22, 90-103. https://doi.org/10.1007/s10567-019-00282-4

Edwards, B., Mullan, K., Katz, I., & Higgins, D. (2014). The Stronger Families in Australia (SFIA) study: Phase 2 (Research Report No. 29). Melbourne: Australian Institute of Family Studies. https://aifs.gov.au/publications/stronger-families-australia-sfia-study-phase-2

Eisenstadt, N. (2011). Providing a sure start: How government discovered early childhood. Bristol, Policy Press.

Fagan, A. A., Hawkins, J. D., Catalano, R. F., & Farrington, D. P. (2019). Communities that care: Building community engagement and capacity to prevent youth behavior problems. New York, Oxford University Press.

Homel, R., Freiberg, K., & Branch, S. (2015). CREATE-ing capacity to take developmental crime prevention to scale: A community-based approach within a national framework. Australian and New Zealand Journal of Criminology 48(3): 367-385.

Homel, J., Homel, R., McGee, T. R., Zardo, P., Branch, S., Freiberg, K., Manning, M., & Wong, G. (2021). Evaluation of a place-based collective impact initiative through cross-sectoral data linkage. Australian Journal of Social Issues, 56(2), 301-318.

Kania, J., & Kramer, M. (2011). Collective impact. Stanford Social Innovation Review, Winter, 36-41.

Katz, E., Rose, J., Low-Choy, S., & Homel, S. (2021, forthcoming). The importance of place-based services on child maltreatment: Evaluation through big data linkage and analytics. In K. Maguire-Jack & C. Katz (Eds.), Neighborhoods, communities and child maltreatment. New York: Springer.

Krakouer, J., Wu Tan, W., & Parolini, A. (2021). Who is analysing what? The opportunities, risks and implications of using predictive risk modelling with Indigenous Australians in child protection: A scoping review. Australian Journal of Social Issues, 56(2), 173-197. https://doi.org/10.1002/ajs4.155

Melhuish, E., Belsky, J., & Barnes, J. (2010). Evaluation and value of Sure Start. British Medical Journal, 95(3), 159-161. https://doi.org/10.1136/adc.2009.161018

Muir, K., Katz, I., Purcal, C., Patulny, R., Flaxman, S., Abello, D., . . . Hayes, A. (2009). National evaluation of the Stronger Families and Communities Strategy 2004–2009. Sydney: Social Policy Research Centre, UNSW.

Paulson, R., & Bailey, P. (2018). ChangeFest 2018 Statement: The national celebration of place-based social change. https://changefest.com.au/

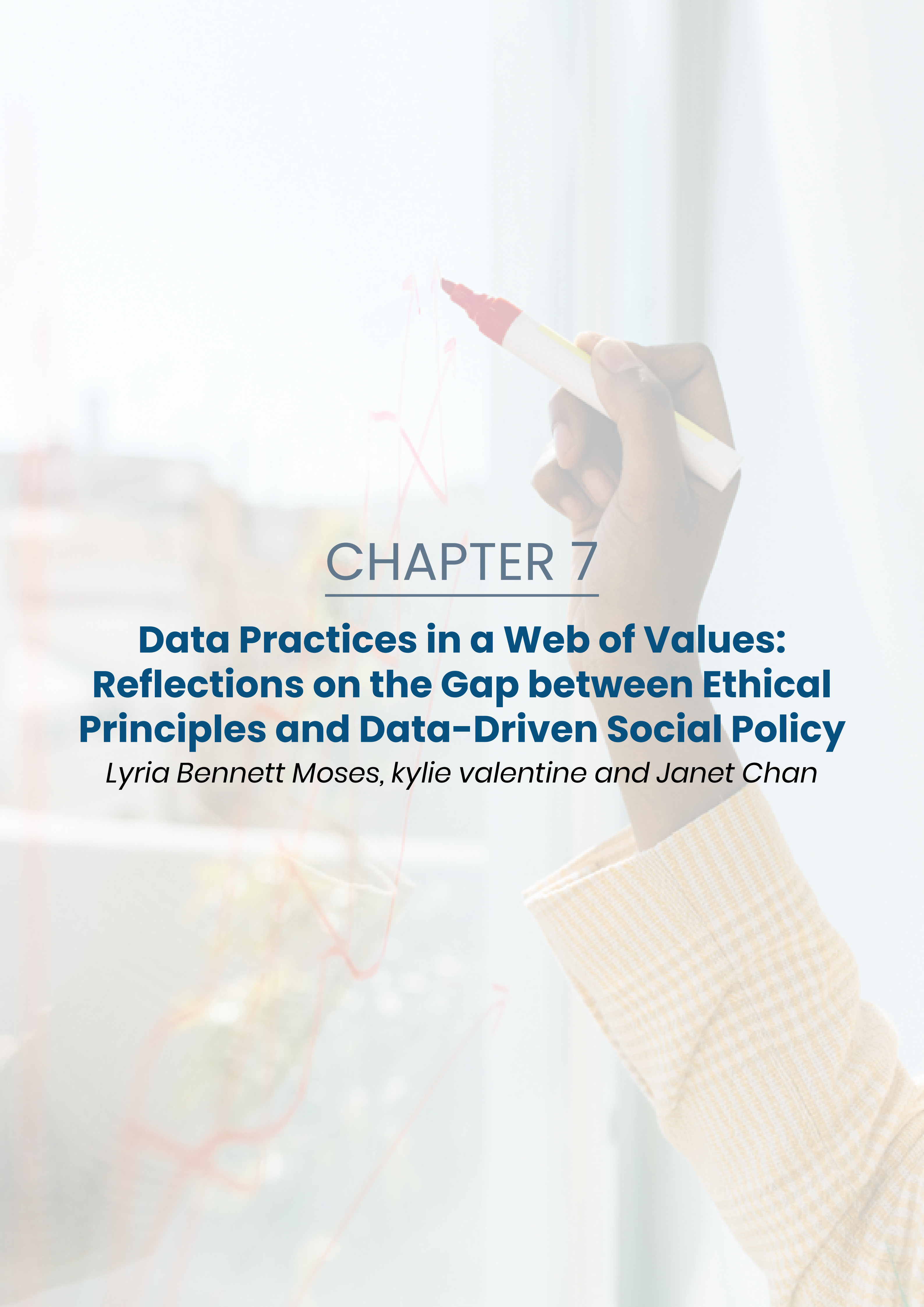
[‡‡] *For example, the South Australian Early Childhood Data Project: https://health.adelaide.edu.au/betterstart/research/*

Preskill, H., Parkhurst, M., & Splansky Juster, J. (2014). Guide to evaluating collective impact: 01 Learning and evaluation in the collective impact context. Collective Impact Forum. https://www.collectiveimpactforum.org

Productivity Commission (2017). Data availability and use (Report No. 82). Canberra: Australian Government.

Wallace, J. (Ed). (2019). The enabling state: Where are we now? Review of policy developments 2013-2018. UK: Carnegie UK Trust. https://apo.org.au/node/252746

World Economic Forum (2019). Data collaboration for the common good: Enabling trust and innovation through public-private partnerships. Geneva: World Economic Forum.



CHAPTER 7

Data Practices in a Web of Values: Reflections on the Gap between Ethical Principles and Data-Driven Social Policy

Lyria Bennett Moses, kylie valentine and Janet Chan

Data Practices in a Web of Values: Reflections on the Gap between Ethical Principles and Data-Driven Social Policy

INTRODUCTION: ETHICAL PRINCIPLES

This chapter examines the ethical dimension of the data-driven social policy initiatives reported in Chapters 2 to 6 and discusses the apparent gap between ethical standards and the realities of policy implementation. Ethics has become a buzzword, as people, governments and international organisations struggle with the implications of processing large volumes of data, using complex computational techniques, to make decisions with high impact on people's lives. In particular, there are a growing number of documents with a variety of labels (framework, statement, charter, declaration, guide, guidelines, principles, among others) that set out how this might be done ethically or in accordance with human rights (Fjeld, 2020: 8-9). Despite the diversity of nomenclature, it is convenient to refer to such documents collectively as setting out "ethical principles." While authored by diverse bodies (international and regional bodies, government, industry, civil society, professional society, academia), ethical principles address similar themes (Jobin et al., 2019).

These ethical principles are not about how humans ought to conduct themselves in general, but rather, are directed towards activities captured under particular labels. For the activities considered here, the most common label is the somewhat misleadingly named "artificial intelligence" or "artificial intelligence systems" (OECD, 2019). Other targets include "the data revolution" (Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014), "autonomous" or "autonomous and intelligent systems" (IEEE, n.d.), "algorithms" or "algorithmic decision-making" (Algo Aware, 2018), and more specific domains such as "robotics" (Engineering and Physical Sciences Research Council, 2010) and "machine learning" (Access Now, 2018). These terms are sometimes used in combination, for example, "algorithms and artificial intelligence"(Commission Nationale Informatique & Libertés, 2017).

Ethical principles operate at a very high level of generality and employ terms that are themselves "essentially contested" (Gallie, 1964), such as fairness, beneficence and accountability. There are various ways in which values articulated in distinct ethical principles might be grouped (Fjeld, 2020: 20-63; Jobin et al., 2019; Guihot & Bennett Moses, 2020: 55-68). Along a single dimension (such as transparency) one might find a range of terms used in ethical principles promulgated by different entities, including intelligibility, interpretability, explainability, contestability, auditability, awareness of misuse—all of which overlap but are not the same (Guihot & Bennett Moses, 2020: 55-68). There may be other important elements too beyond those found in ethical principles (e.g., Henman, 2020).

While there is significant agreement across ethical principles on the themes that are important (albeit using diverse concepts and terminology), there is a significant gap between what is agreed to be important and what occurs in practice. Our objective is to examine the extent to which such high-level values are evident in practice, drawing on the detailed examples presented in this book. We do not provide a full analysis (all of the principles, as articulated in every set of ethical principles, for all of the examples) but instead, focus on what the examples reveal about the extent to which ethical principles have informed the design of these programs or are reflected in their implementation and outcomes. We also use this analysis to reflect back on the ethical principles themselves—in particular, whether they are reflected in practical policy contexts.

A useful framework for our analysis is provided by the Organisation for Economic Co-operation and Development's Principles on artificial intelligence (OECD, 2019), not because they are necessarily the "best" (analysed from an ethical perspective), but because they purport to apply across the jurisdictions in the examples. In particular, they have been adopted by all OECD members, including Australia, New Zealand, and the United States. That is, in assessing examples against the principles, we are not making claims about the success or failure of the examples to live up to ethical ideals, but assessing their operations against the standards agreed to be feasible and acceptable by national governments.

The OECD Principles concern AI systems, defined broadly as:

> *… a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.*
> *(OECD, 2019: clause I, p. 7)*
>
> The OECD Principles are
>
> **1.1 Inclusive growth, sustainable development and well-being**
>
> Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.
>
> **1.2 Human-centred values and fairness**
>
> a. AI actors [defined to include individuals and organisations that deploy or operate AI] should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.
>
> b. To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.
>
> **1.3 Transparency and explainability**
>
> AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:
>
> i. to foster a general understanding of AI systems,
>
> ii. to make stakeholders aware of their interactions with AI systems, including in the workplace,
>
> iii. to enable those affected by an AI system to understand the outcome, and,
>
> iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.
>
> **1.4 Robustness, security and safety**
>
> a. AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.
> b. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.
> c. AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.
>
> **1.5 Accountability**
>
> AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art. (OECD, 2019, clause IV, pp. 7-8)

While these are relatively typical of principles directed to computational techniques, other principles focus on data. A typical list for such principles would include data quality, data openness and useability, data protection, and accountability of analysis as well as human rights and broader issues that arise across the board (Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014).

An interesting hybrid incorporating computational techniques and data practices is the *Algorithm Charter for Aotearoa New Zealand* (New Zealand Government and Stats NZ, 2020), designed to ensure public confidence in how government agencies use "algorithms" as well as transparency and accountability in the use of data. The definition of "algorithm" is broad and includes "less advanced techniques such as regression models and decision trees," and the charter recognises that harm can result from even simple algorithms. The focus is on risk of harm (taking account of likelihood and impact) rather than technique. Agencies signing on to the charter commit to:

> **TRANSPARENCY**
>
> Maintain transparency by clearly explaining how decisions are informed by algorithms. This may include:
>
> - Plain English documentation of the algorithm
> - Making information about the data and processes available (unless a lawful restriction prevents this)
> - Publishing information about how data are collected, secured and stored.
>
> **PARTNERSHIP**
>
> Deliver clear public benefit through Treaty commitments by:
>
> - Embedding a Te Ao Māori perspective in the development and use of algorithms consistent with the principles of the Treaty of Waitangi.
>
> **PEOPLE**
>
> Focus on people by:
>
> - Identifying and actively engaging with people, communities and groups who have an interest in algorithms, and consulting with those impacted by their use.
>
> **DATA**
>
> Make sure data is fit for purpose by:
>
> - Understanding its limitations
> - Identifying and managing bias.
>
> **PRIVACY, ETHICS AND HUMAN RIGHTS**
>
> Ensure that privacy, ethics and human rights are safeguarded by:
>
> - Regularly peer reviewing algorithms to assess for unintended consequences and act on this information.
>
> **HUMAN OVERSIGHT**
>
> Retain human oversight by:
>
> - Nominating a point of contact for public inquiries about algorithms
> - Providing a channel for challenging or appealing of decisions informed by algorithms
> - Clearly explaining the role of humans in decisions informed by algorithms. (New Zealand Government and Stats NZ, 2020)

While these principles purport to provide a pathway for ethical conduct in the intersecting spaces of big data, data processing, artificial intelligence and autonomous decision-making, there are few mechanisms to ensure that they are put into practice. It is easier to agree that transparency or fairness are essential for ethical delivery of programs than to communicate methods or minimise unfair disparate impact on marginalised groups. As the examples in this collection illustrate, there is a significant gap between ethical aspirations and current reality. This raises not only issues of willingness to implement principles, but also the usefulness of the principles themselves in guiding action. However, it is important to note that the programs discussed in this collection pre-date (at least in part) formal adoption of ethical principles. It would thus be worth comparing our findings with new projects going forward.

This chapter seeks to describe and understand that gap in the context of the examples described in Chapters 2 to 6. It begins with a discussion of the problem of scope—the fact that real-world government programs align poorly with the scope and mission of the OECD Principles. More broadly formulated principles, such as the Algorithm Charter for Aotearoa New Zealand (New Zealand Government and Stats NZ, 2020) have greater potential. It then describes, for the examples, the ways in which they fall short of the aspirations set out in the OECD Principles (and similar statements). Subsequently, we explain how such failures can be understood against trends in the conception and implementation of policy programs, such as those exemplified by the examples in this book, and what this might say about the formulation of the principles themselves.

THE SCOPE OF ETHICS FOR DATA PRACTICES

In philosophy, ethics or moral philosophy explores values relating to human conduct—in other words, how do we decide whether a person ought to act in a particular way? It is complex and contested: we could make decisions based on actual or predicted outcomes, on motive and internal reasoning, on alignment with a set of norms (such as human rights), or on some other basis or combination of bases. Studying moral philosophy takes place among these debates about how we ought to act and make decisions about how to conduct ourselves. However, most ethical principles articulated for subject matter, such as artificial intelligence or big data are not exclusively addressed to moral reasoning by humans—they also demand that systems have certain features, such as fairness, transparency and beneficence.

Ethics is not the same thing as law. When an ethical norm demands “fairness,” this is separate from legal rules that might insist on non-discrimination against particular groups in the context of particular decisions and actions. Non-compliance with legal rules typically has consequences such as paying damages or fines, and court orders requiring decisions to be re-made in line with the rules. Non-compliance with ethical principles can result in negative publicity or consumer boycotts, but there is no state enforcement mechanism, and generally no mechanisms for audit, investigation or certification. Technical standards (such as those issued by the International Standards Organisation) generally provide more specific guidance than ethical standards and may allow for certification of conformance against requirements contained in the standard.

As noted above, ethical principles such as those formulated by the OECD only apply within specified domains. Not only are a variety of different terms used to describe the scope of application, but different principles define the same term differently. For example, one can contrast the OECD’s definition of AI systems with the definition of artificial intelligence of the Council of Europe Commissioner for Human Rights (“set of sciences, theories and techniques dedicated to improving the ability of machines to do things requiring intelligence”, Council of Europe 2019:5). It is thus worth considering the extent to which the examples fall within the domain of different ethical principles that, from a jurisdictional perspective, could otherwise apply.

The New Zealand example (Chapter 2) involved actuarial analysis based on linking diverse administrative data. This was described in the chapter as involving “calculation of long-term projections of costs and liabilities, estimating the ‘risk’ that an individual will make a claim or use a benefit.” If this were simply used for budget forecasting, the only real question would be accuracy. Instead, it is used to make decisions about “investment” in potential groups of beneficiaries, including children at risk. However, actuarial analysis is typically led by certified human actuaries and other skilled personnel as opposed to a “machine-based system.” Thus, while computational tasks would be undertaken by computers, this is unlikely to qualify as “artificial intelligence.” While earlier stages of this initiative pre-date New Zealand’s Algorithm Charter, the program now falls within New Zealand’s broad definition of “algorithms,” given the potential for risk. Indeed, the Social Wellbeing Agency is on the list of founding signatories to the charter. Principles focused on data practices, such as the Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014), would also apply.

Value-Added Models for evaluation of teaching performance (Chapter 3) are based on linear models that factor in variables such as current and prior test scores, student characteristics and classroom characteristics. The volume, variety and precision of data employed in using these models on a national scale to assess teachers and students brings this example within the scope of ethical principles that focus on data practices, including the data revolution or big data. However, whether the use of linear models constitutes artificial intelligence is contingent. For the purposes of the OECD definition, the question would be whether this is a "machine-based system" or whether machines are simply performing calculations as directed by humans. The calculations required could be done either way. Principles focused on data practices would apply.

Robodebt (Chapter 4) involved linking data between two government agencies, automation of calculations and the dispatch of debt letters. As in Chapter 5, the data linking process most likely involves artificial intelligence, although that aspect of the process was relatively uncontroversial. The calculations, which problematically assumed that fortnightly income was proportionate to annual income, were performed using simpler programs. It likely does not fit easily into the OECD definition of AI system because the system is not given "human-defined objectives," rather, it simply followed a series of pre-programmed procedural steps that led to an output (being a "decision" to issue a debt letter).

The strategic collection, linking and analysis of diverse administrative data to improve the wellbeing of vulnerable children is discussed in Chapter 5 and Chapter 6, the latter focusing specifically on place-based initiatives. Ethical principles that focus on data practices clearly apply to the data linkage projects discussed there. Most data linkage tools are probabilistic and rely on machine learning techniques, and thus would involve artificial intelligence, according to most definitions. However, not all programs would rely on such techniques.

Ethical principles that focus on data governance cover more of the examples than those focusing on artificial intelligence and autonomous systems. Yet, as will be demonstrated below, attention to requirements such as human-centred values and fairness, transparency and explainability, robustness and accountability, would all have been of benefit. This suggests that ethical principles for artificial intelligence are conceived too narrowly or are using the wrong buzzwords. The ethical principles formulated by the OECD (and others) should apply to all government programs involving data-driven inferencing that leads to different treatment. This seems to be the objective of the Algorithm Charter for Aotearoa New Zealand (New Zealand Government and Stats NZ, 2020), whose scope is aligned with risk rather than sophistication of technique. Thus, if we want ethical principles to cover the kinds of programs considered in this book, which we believe would be wise, a broad scope that is not limited along the domain of technical sophistication is essential. This recognises that ethical challenges are not limited to artificial intelligence (or any other category) but may rather be exacerbated by particular data and machine learning practices (Henman, 2020).

ETHICAL ALIGNMENT: PRINCIPLES VERSUS REALITY

Chapters 2 to 6 provide some empirical evidence of how "big" or "smart" data is used for social policy in practice. This section examines the extent to which the OECD Principles are evident in the design, implementation and outcomes of these programs. In particular, we highlight examples where significant departures are evident.

Inclusive growth, sustainable development and wellbeing

The first OECD Principle is difficult to assess, as it depends on one's views of the policy goals of each program and their relationship with other policies and values. All of the programs discussed in this book had goals that could be described as seeking beneficial outcomes—identifying children at risk, improving teaching and hence educational outcomes, meeting local needs, and ensuring fair and accurate distribution of government welfare. However, some of these programs had net negative effects—Robodebt (Chapter 4) had a negative impact on the wellbeing of those receiving debt letters, whether or not the calculations were accurate, and failed in its intended purpose of raising government revenue. Value-Added Models for evaluation of teaching performance (Chapter 3) also had negative outcomes that countered inclusivity and wellbeing because teachers were able to game the system in ways that increased disadvantage for high-need students. Satisfaction of the first principle is thus inevitably tied up with policy preferences and will rarely change a policy program, given that (presumably) policies are implemented because leaders believe they will have a positive impact.

Human-centred values and fairness

The second principle implies that data-driven systems should “respect the rule of law, human rights and democratic values,” including values such as “freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally labour rights” (OECD, 2019). This applies to both the design of the systems and their implementation, which suggests that such systems should include mechanisms and safeguards to ensure that those values are protected.

The NZ social investment approach (Chapter 2) involves the use of actuarial analysis of linked government administrative data to identify groups in need and develop and deliver strategies for reducing reliance on welfare over the long term. Its design was primarily focused on the value of “rigorous and evidence-based” analysis (Chapter 2), as well as collaboration between agencies, although such ideals were not necessarily realised in practice, given the lack of shared understanding about the aims of the program, the variable quality of data linkage/data quality, the inadequacy of technical competency in data analysis and the lack of incentives for cross-agency collaboration. There was no apparent attention to values such as the rule of law and human rights. Fairness, privacy, data protection, social justice and dignity do not seem to have been explicit design requirements, nor were there mechanisms to safeguard or protect these values in the implementation. The lack of policy evaluation and external peer review meant that it was impossible to ascertain whether the program had achieved outcomes that were consistent with fairness, diversity and non-discrimination. In contrast, the example has revealed concerns regarding risks to human rights, dignity and autonomy, privacy and data protection, as a result of its failure to secure the social licence to use data on Indigenous people, and the failure to engage with Indigenous people regarding control over the collection and analysis of their data. The risk of increased stigmatisation of Indigenous people and other vulnerable groups was raised as a critique of the system, and was also mentioned in Chapters 5 and 6.

“Value-Added Models in Education Performance Management” (Chapter 3) reviews programs aimed at improving the performance of the education system by using statistical models to assess the performance of teachers and/or schools. The model was primarily based on the test scores of students in standardised educational tests, usually controlling for other factors such as students’ prior test scores, demographics, and other student-level variables. There was no explicit reference to values such as human rights, freedom, dignity, autonomy, privacy, etc. in the design or implementation of the program, although in the United States, the programs were based on legislation (RTT, ESS 2016), and had been subject to 15 lawsuits by teacher unions and others, on a variety of grounds related to the values discussed in this section. Questions were also raised about the validity of models and their use in decisions such as teacher employment. There was concern that the metrics assigned to particular teachers, and thus decisions made about them, may have been unfair for a variety of reasons that were not factored into the statistical analysis. The outcomes of these programs were problematic in terms of fairness and the human-centred values of interest in this discussion: the programs led to unintended consequences such as the inability of some school districts to retain teachers, teachers refusing to collaborate with other teachers, and teachers seeking positions where their “performance” would not be affected by factors such as high-need students. The issue thus goes beyond fairness to teachers to include fairness to communities and students.

“Big Data and Government Services” (Chapter 4) examines the Robodebt experience, as well as the matching of administrative data across agencies more generally. The design of the Robodebt program was aimed at identifying discrepancies in income (between the Australian Tax Office (ATO) and Centrelink) in order to recover “debt” and reduce welfare expenditure. Issues such as the rule of law, human rights and democratic values were not part of the design, nor did the program include mechanisms or safeguards to protect privacy, dignity, or fairness in its implementation. In fact, when benefit recipients attempted to challenge a “debt” notice, they encountered long telephone waits. The program was subject to a number of legal challenges, although this was not part of the safeguard built into the system. The outcomes of the program suggest an absence of fairness, social justice, and due process: for example, the program was based on erroneous assumptions about reliability of inferences that could be drawn from annual income data; welfare recipients with disability or disadvantage (such as no access to computers or lack of computer literacy) were unfairly targeted; the approach was punitive, shifting the onus of proof on recipients to prove their innocence; and many debt letters sent to vulnerable people were erroneous. From an individual perspective, receiving an erroneous debt letter is “unfair”—particularly when the onus was on them to provide the evidence to identify the error. From a broader perspective, the program was unfair in that it targeted vulnerable groups (as opposed to focusing revenue collection on wealthy individuals evading taxes). More generally, the example suggests that data matching requires greater attention to issues such as the appropriateness of the model, data quality, social licence, and data protection.

"Strategic Data Linkage to Improve the Wellbeing of Vulnerable Children" (Chapter 5) explores the promise of linking and analysing administrative data from multiple providers over time to improve service outcomes for children. Since this study is not based on an evaluation of a specific program, its relevance relates to how such programs could be designed, implemented and evaluated to maximise human-centred values and fairness. Issues such as social licence, consent from data subjects, ethics protocols, confidentiality (especially in small communities), and cultural acceptance are raised as important values to incorporate in the design and implementation of such programs. The chapter on place-based initiatives (Chapter 6) specifically mentions risks of stigmatisation as a result of analysing biased data. The example also raises technical issues such as data quality, data security, linkage accuracy, resources and skills, and data interpretation that may affect ethical issues such as fairness, due process and misleading outcomes.

Transparency and explainability

The next group of values specifies that data-driven systems should "commit to transparency and responsible disclosure," including the provision of "meaningful information" to enable stakeholders and those affected by the system to become aware of and understand the system and its outcome, and those adversely impacted to challenge the outcome (OECD, 2019). To be "meaningful," the information should be "plain and easy-to-understand" and include details such as "the factors, and the logic that served as the basis for the prediction, recommendation or decision" (OECD, 2019). Transparency as a value is closely entwined with the rule of law value of accountability (Zalnieriute et al., 2019; Henman, 2020), may be a condition for a social licence to operate even beyond legal requirements (Carter et al., 2015), and has a complex relationship with trust (O'Neill, 2002).

In general, none of the examples in Chapters 2 to 6 shows that transparency and explainability were features in the design or implementation of the programs being examined, although the lack of transparency was raised explicitly or implicitly as a problem in some cases. For example, the NZ social investment approach (Chapter 2) criticises the lack of transparency in how the data should be analysed and how outcomes were to be defined and measured, and the proprietary nature of the models used. Similarly, "Value-Added Models in Education Performance Management" (Chapter 3) report that the lack of transparency, the complexity and proprietary nature of algorithms and results, and the lack of audit processes for correction of erroneous information can impact on due process as a basis for lawsuits against the program. The need for better research on implementation as well as mechanisms for improving the understandability of decisions, allowing contestations of the accuracy of the data/results, and avenues for redress are highlighted as significant flaws of the program. Chapter 4 refers to transparency, not only in relation to Robodebt, but also in the context of the lawfulness of data analytics more broadly. "Strategic Data Linkage to Improve the Wellbeing of Vulnerable Children" (Chapter 5) refers to transparency (and, in particular, openness with vulnerable groups about what data is collected, how it is used, and why it is necessary), in connection with privacy and trust. In this context, transparency is one aspect, identified as important, but not analysed specifically.

Robustness including traceability re datasets

The principle of "robustness, safety and security" sets out three criteria for data-driven systems to meet. First, systems should be "robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk." Second, this requires traceability of datasets, processes and decisions, "to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art." Third, systems should have a comprehensive risk management approach for "each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias."

Evidence-based policy demands robust data, and both policy practitioners and scholars have been concerned for decades with the robustness of the data used for social policy. The evidence from the examples in this book indicates that significant gaps and constraints remain, despite this long attention, and, in fact, are probably worsening as data systems and technologies evolve. The consequences of these gaps are that many systems do not meet the OECD criteria, although this is not an intended effect of approaches to managing social policy data—they are not in opposition to the OECD Principles—instead, it is a function of the reliance of social policy on multiple datasets, many involving a skewed sample, collected for different purposes, and often in resource-constrained environments. Single datasets, such as the census, are essential for tracking key outcomes in populations over time. Yet single datasets are not sufficient in many cases. Increasingly, social policy interventions are directed at health, wellbeing and participation outcomes of smaller groups of people, whose effects are multidimensional. With respect to data linkage in relation to vulnerable children (Chapter 5) and place-based initiatives (Chapter 6), the groups of people may be smaller, while variables are likely diverse, coming from linked data held by different agencies. If there are too many variables for a small dataset, robustness is reduced. The challenge may be compounded due to the need for data linkage, which, as the examples show, is often difficult to achieve, for technical reasons, and for other reasons, such as perceived or real barriers to information sharing between agencies and with the private sector. Moreover, the quality of individual datasets is often low, either because it is collected by service provider organisations whose core business is service delivery and not data collection, or because the data does not address outcomes or significant variables directly, or both. There are additional issues for place-based initiatives, discussed in Chapter 6, with people moving geographically and thus entering different programs over time, and with evaluation challenges in the context of different layers of programs in different places, with little cross-comparison and no true control group.

The chapters describe gaps in infrastructural and workforce capacity, with interview participants who are stakeholders from policy, practice, and research providing many examples of both. All of these have implications for the robustness and security of data. Importantly, capacity is needed in both technical terms, to manage the storage and security infrastructure, and analytic terms. As Chapter 3 notes, this includes capability and evidence related to implementation. These are different skills and require different workforces, all in addition to the need for data-related skills for effective, secure data linkage and analysis.

For the most part, then, the move to data linkage and greater use of data was described with enthusiasm by stakeholders, who nonetheless also described the significant need for improvement in data quality, due mostly to limitations in resources and capacity. The exception to this is Robodebt (Chapter 4), which was not a failure of capacity but a failure to use existing capacity. In 2020, the Commonwealth Government settled a class action, apparently the biggest in Australian history, at a cost of $1.2 billion, comprised of refunds of $721 million to 373,000 people, $112 million in compensation and $398 million in cancelled debts (Whiteford, 2021). The class action reflected a breach of every one of the OECD Principles, but the significance of data robustness to both the scheme and its failure is stark. Robodebt, as noted above, imputed fortnightly income from ATO data on annual income without allowing for any fluctuations throughout the year. This was not an error, however, but a design choice.

Accountability

In the OECD Principles, accountability operates vertically across other principles, with AI actors made accountable for proper functioning as well as compliance with each earlier principle. Accountability refers to “the right of the account-holder to investigate and scrutinise the actions of the agent by seeking information and explanations and the right to impose remedies and sanctions” (Mulgan, 2003: 10).

Understanding a requirement for accountability requires answering five questions: who, for what, to whom, how and according to what criteria (Mashaw, 2006: 117-18). According to the OECD Principles, the obligation is owed by “AI actors,” defined as “those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI.” In the context of government programs, this would include the relevant government departments as well as outsourced providers of systems. These actors are accountable for proper functioning and compliance with other principles, the criteria for which have been discussed above. The principles are silent on questions of to whom and how, which is left to the national level.

Accountability is a fundamental value requiring government to answer for its actions and decisions, and encompasses at least lawfulness, fairness, transparency, and rationality (Boughey & Weeks, 2018; Mulgan, 2003). Its inclusion or exclusion in ethical principles does not alter its crucial role in the rule of law, at least as far as government action is concerned. Many of the failures described in relation to Chapters 2 to 4 are failures of accountability as well as of fairness, transparency, robustness, and so forth. In other words, problems arise because of the absence of sufficient measures to hold agencies to account for the use of systems.

This is particularly apparent in the Robodebt case (Chapter 4). Prior to implementation of the scheme, staff were responsible for investigating non-compliance based on evidence obtained in the individual case, often using information gathering powers to request relevant data from employers. As well as the internal hierarchy to which such staff were accountable, the pace of processing was slow, and appeals could be made to the Administrative Appeals Tribunal. Following implementation of the scheme, accountability no longer existed along internal human hierarchies—the system simply issued a debt letter with instructions to pay, with no internal human checking processes. Explanations to those affected were difficult to procure (Victoria Legal Aid, 2019). While internal appeals and appeals to the Administrative Appeals Tribunal remained possible after the letter was issued, the rate at which the system issued erroneous letters made it difficult for these mechanisms to keep up. In particular, there were long telephone waiting times and staff were often unable to assist beyond pointing people to the confusing online platform (Commonwealth Ombudsman, 2017). Further, the process required individuals to provide historical evidence from prior to the required retention period for income evidence.

Many concerns expressed in Chapter 4 about data linking come down to the question of accountability. The lack of trust in government (data sold to business, lack of security measures, sharing without checks and balances, data misuse) corresponds with insufficient accountability measures. With proper processes in place, including independent oversight of those processes, the risks of such things occurring can be minimised and the social licence for data-sharing activities may be enhanced. Eventually, the Robodebt program was held to account, primarily through activism and public journalism, but also threat of legal action, resulting in its demise. It is not clear whether the lessons of this failed experiment will be learnt (so that future automation takes the need for accountability of the program into account). Without this, social licence for similar programs is unlikely.

In the New Zealand example on social investment (Chapter 2), some of the concerns raised by research participants relate to accountability. For example, Participant A raised the issue of outsourcing to financial services firms as reducing transparency, but this can also impact on accountability. If the government relies on non-transparent analysis for its decision making, how can it give an account of “investment” decisions that advantage some groups over others or that classify some groups as “risky”? Even though social investment is about groups rather than individuals, differential assumptions or treatment should be explained and justified both to those affected and the broader public. While the techniques were not “artificial intelligence,” “algorithms” were likely involved, as it is unlikely the processing of data was entirely by hand. Had New Zealand’s Algorithm Charter been in effect at the time and adopted by relevant agencies, the requirement to “retain human oversight” would not have been enough to ensure accountability. This is because the lack of transparency and hence accountability was related to outsourcing to humans in the private sector (relying on proprietary processes) rather than machines. The problem was not the lack of a human role in decisions, but a question of whether the humans making decisions were accountable to the public. Nevertheless, a clear government contact authorised to answer questions would have been useful, provided such answers were not restricted by the terms of outsourcing. An additional issue in this example (and others) is whether there ought to have been accountability specifically to specific groups such as Indigenous communities as well as the public more broadly, essential for a social licence to use their data.

Strategic data linkage (Chapter 5) also involves questions of accountability to segments of the public rather than the broader public, particularly for placed-based initiatives (Chapter 6). Like the NZ example (Chapter 2), it involves decisions to benefit groups rather than individuals, sometimes based on a single variable (place of residence, indigeneity). The lack of fully realised examples in the study makes it difficult to analyse accountability, in particular to local communities, in practice. However, it echoes Chapter 2’s point regarding Indigenous involvement in governance, raising also broader questions of co-creation of programs with intended beneficiaries and “working with and not on people” (Chapter 6). The accountability processes for data sharing would also change if the *Data Availability and Transparency Bill*, discussed in that study, is passed.

The Data Availability and Transparency Bill in Australia promises more accountability—both in decisions about what data is shared and with respect to the obligations. The Bill would likely reduce cultural reticence to share (Chapter 4) and associated difficulty obtaining access to administrative datasets (Chapter 5), by establishing a clear statutory mechanism for data reuse. While organisations with custody of particular datasets will retain an option not to share under the scheme, informing government policy and programs is a “data-sharing purpose” encouraged by the Bill (clause 15(1)(b)). However, what is particularly useful is the proposed accountability measures where sharing takes place. These include Chapter 4 and Part 6.2 of the Bill, dealing with requirements related to data sharing through systems for accreditation, complaints, assessments and investigations, regulatory powers and enforcement, and appeals of decisions. All of these require those sharing and receiving data under the scheme to provide accounts of their actions to the National Data Commissioner, as well as, in some cases, courts and tribunals.

“Value-Added Models in Education Performance Management” (Chapter 3) involved a program introduced for school and teacher accountability, and yet failed to be accountable in its own operation, that is, the accountability of the managers of the program themselves. As in the New Zealand example, the most significant hurdle to accountability is lack of transparency due to proprietary methods. The implications of a lack of transparency are, however, more significant than for the New Zealand example because VAMs are used to make operational decisions with very real impacts on individuals’ livelihoods and careers as well as local school communities. In this context, accounts should be required to those affected with opportunities for appeal and/or oversight. This links with the due process demanded in lawsuits brought in relation to the program.

Like other government programs, accountability should be a requirement independent of technique. Yet this program is not the target of ethical principles focusing on artificial intelligence. While the OECD’s ethical principles are thus out of scope, the involvement of the OECD in promoting VAMs without demanding accountability around their use suggests that priorities of different working groups may not align, and highlights the limitations of linking broad objectives (such as accountability) to specific domains (such as AI). New Zealand’s charter, being more broadly framed, would apply if these programs were implemented in that country by an agency adopting it. That would introduce requirements for explanations, contestability, and human oversight that are otherwise lacking in the use of VAMs in many jurisdictions.

Accountability is an important principle, best considered from a broader perspective than merely “artificial intelligence.” What needs to be accountable are socio-technical systems employed by governments and other powerful actors where such systems risk unlawful, unfair, or irrational outcomes. There may be different specific mechanisms to achieve accountability depending on the context—appeal processes, the timing and form of explanations; the relationship between internal and external accountability mechanisms and so forth will hinge on context, including local institutional frameworks. But the scope of an accountability principle should be framed around who is accountable (which should include government actors) as well as matters such as to whom, how, and in what circumstances, rather than the nature, extent or sophistication of the technology employed.

THE POLICY CONTEXT: TRENDS IN THE REAL WORLD

The idealism expressed in ethical principles about the need for fairness, transparency, accountability and so forth needs to be understood in the context of how policy programs are formulated and implemented. In this section, we discuss how techno-optimism associated with ideas around big data and artificial intelligence intersects with longstanding technical and ethical concerns. This highlights that new ethical principles, such as the OECD Principles, targeting new techniques, are unlikely to make an impact unless understood in a broader historical and technical frame.

As policy frameworks and design are increasingly influenced by automated decision making and inferences drawn from big data analysis using artificial intelligence techniques, distinct and sometimes contradictory trends emerge. First, this policy landscape is characterised by technoscientific optimism that better uses of data can address longstanding technical and ethical problems, improve policy implementation, and lead to improved wellbeing. Second, it is equally characterised by cautions and anxieties, including limitations on availability of data, organisational technical capacity, requirements of privacy law, and the growing importance of ethics. The recent initiatives MyHealth, a patient electronic health record, and COVIDSafe, an app for contact tracing, exemplify both of these trends. In both cases, advocates for the initiatives, especially the government agencies that developed them and relied on their widespread take-up for them to be effective, emphasised their potential benefits and safety. On the other hand, public commentary and a lower than anticipated take-up rate suggested scepticism and reluctance, expressed in terms of privacy concerns and a lack of confidence in the purported capacity of the initiatives to perform as intended (Bennett Moses et al., 2020; Leins et al., 2020; Lupton, 2019; Pang et al., 2020).

Big data's use in social policy formation is then a context in which pre-existing, and often longstanding, social and ethical concerns remain salient, and are accompanied by new concerns. The examples here explicate both. Stakeholders from policy and research advocate for the anticipated advantages and benefits of new forms of data, and techniques for processing data being developed at scale. At the same time, the potential harms and risks of policy have not been addressed by changes in technology, and in some cases have exacerbated them (Eubanks, 2018; Obermeyer et al., 2019). Public concerns about emerging technologies (Macnaghten et al., 2019), and about the safety and privacy of the personal data held by institutions, have acted as barriers to the take-up of data-driven initiatives such as MyHealth and COVIDSafe (Lupton, 2019). This can then contribute, in turn, to the ineffectiveness of these initiatives, which rely on data to work. Policy formation remains a complicated and contested arena, in which different interests and different forms of evidence are important. Policy formation related to big data and artificial intelligence is particularly fraught in these areas because of the rapidly changing nature of the field, and because the stakes are very high in terms of social and ethical consequences (Crawford et al., 2016).

Data linkage is central to Chapters 4 to 6 in this book. Data linkage is not new, although the technical capacities have increased over time. The potential utility of data linkage is then in its expansion: it is potentially more efficient and achievable at a much larger scale. Chapters 5 and 6 set out the policy imperatives to facilitate greater use of linked data, that is, to identify whether and how policies intended to improve the outcomes of disadvantaged communities are doing so. Increased use of data, including new uses of existing data, has the potential to contribute to the evidence base on policy implementation, efficiency and effectiveness. Stakeholders interviewed for Chapters 2 and 4 also discussed the benefits of increasing the use of linked data. The relatively new policy initiatives of "social investment" and Value-Added Models and the relatively novel application of these techniques to new social policy areas rely on the linking of multiple data systems to realise their anticipated benefits.

To realise these benefits will take resources, and the examples in this book provide extensive, consistent information about resource needs and gaps. These include resources to build hardware and infrastructure, data systems, technical and workforce capacity, and data quality. In New Zealand, the key resource of the social investment approach has been the Integrated Data Infrastructure dataset, described by stakeholders as "a great resource" that Australia should emulate, but with important limitations to do with infrastructure capability (participants state that the interface "drops out" or "collapses" regularly) and data quality. In Australia, stakeholders reporting on the potential benefits of linked data for case-based studies, and for social policy more broadly, identified significant needs in technical and workforce development, including interoperable data systems and the skills and expertise required to facilitate data linkage and ensure management of data storage, data risks and sensitivities.

However, increasing the scale of data use in social policy, and its forms of data collection and management, is not a straightforward question of resource allocation and technical development. Concerns relating to social licence, trust and privacy, which are increasingly discussed in the public arena and detailed in the examples in this book, represent an amplification and extension of longstanding concerns about privacy and the ethics of corporate and government uses of citizen data (Newman et al., 2020; Rieder & Simon, 2016; Sadowski, 2020; Zuboff, 2019). In addition, the examples also illuminate new challenges which are specific to the current era. These are not necessarily to do with the novelty of big data initiatives, but their use in new ways. Social investment refers to the use of actuarial techniques to target groups of people identified as at high risk of future fiscal liability for program intervention. As Kalucza and Baxter note in Chapter 2, this type of actuarial analysis is usually employed in insurance settings. It is not new or a product of big data, but its application to social policy is, and "links individual-level data from different government agencies, including tax, health, education and social services, with data from the 2013 and 2018 censuses and other national surveys" (see Chapter 2, p. 24). Robodebt, as has been described elsewhere in this chapter, was not a new process, but the removal of human checking and validation from an existing process. Value-Added Models are the product of advances in technology and statistical modelling and their application to students' performance in testing.

The concerns with these initiatives are then not only to do with the OECD Principles of inclusive growth, sustainable development and wellbeing, human-centred values and fairness; transparency and explainability; robustness, security and safety; and accountability—although these principles are certainly relevant. The initiatives may also amount to a kind of category error, in which data-driven techniques are used in ways that may be technically correct but that are based on problematic premises. Value-Added Models, for example, are compared in Chapter 3 to clinical trials, in which the differences between efficacy and effectiveness are crucial, not only in the lab but in implementation. The methodological kill-shot sought by advocates of VAM, as noted in Chapter 3, which ensures robustness of data, is irrelevant if the outcomes generated in the ideal context of research settings (analogous to a closed clinical setting) are not those of messy real-world conditions in schools and communities. The social investment state has been subject to extensive critique for its focus on a very narrow range of outcomes, that is, future welfare payment. In NZ, this range of outcomes was expanded before the social investment approach became less prominent with a change of government; however, the basis of the selection of cohorts, and the foundation of the approach, is its precision. The narrowness of outcomes, future benefits receipt, is both what enables the approach and a severely limiting constraint on its effectiveness.

These examples are thus not merely examples of ethical failure (measured against the OECD Principles), they also exemplify longstanding problems in the conception, framing and implementation of social programs. The history of such problems lies beyond newer techniques and buzz words (big data, artificial intelligence, and so forth). Ethical principles framed in terms of such techniques are thus likely to be insufficient as a solution.

CONCLUSION—WHO CARES?

Just as techno-optimism infects the use of data in the policy context, so too, optimism around the likely impact of formulating ethical principles affects our response. Ethical principles are formulated around domains (such as artificial intelligence) that align only partially with government programs to which they might be relevant. Those domains also fail to capture the more extended history of ethical failures in the delivery of government programs. In addition, there are few mechanisms for translating high-minded but essentially contested concepts such as “fairness” into the delivery of programs.

Despite the goals of ethical principles such as those formulated by the OECD in relation to artificial intelligence, the examples in this book highlight negative impacts on marginalised groups and groups with low social power. This is because of the failure in each case to meet the principles, despite the government’s formal agreement to them, is partly a question of timing but may also be in part because the principles themselves are insufficiently attentive to existing social inequalities and the risks of AI in exacerbating them. The first principle articulates an aspiration to increase inclusion of underrepresented populations, but the examples, especially Chapter 4, highlight the impact of existing patterns of stigma and marginalisation in the unfair application of policies, and that this works in exclusionary rather than inclusionary ways. The desire to analyse data for the purposes such as “to predict welfare risk and evaluate the outcomes of policy changes” (Chapter 2) generates its own risks. The causes can be found in a failure to embed fairness, transparency, robustness, and accountability in program design. This, in turn, can be linked to variations in data competency and skills (Chapters 2 and 5), poor resourcing, governance and infrastructure (Chapters 2 and 4), lack of social licence (Chapters 2 and 5), misaligned political priorities (Chapter 2), insufficient safeguards and misplaced onus of proof where decisions affect individuals (Chapters 3 and 4), and opacity of algorithms and processes (Chapters 2 to 5). In other words, there are technical, cultural and political barriers to the “ethical” use of data and analytics in government programs. By exploring real examples of policy failure in a broad historical and technical context, we can hopefully find solutions that go beyond technically contingent and vague expressions of ethical principles.

REFERENCES

Access Now (2018, May 16). The Toronto Declaration. Access Now. https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/

Algo Aware (2018). State-of-the-art report: Algorithmic decision-making. https://actuary.eu/wp-content/uploads/2019/02/AlgoAware-State-of-the-Art-Report.pdf

Bennett Moses, L., Churches, G., Zalnieriute, M., Byrnes, A. C., Scully, J. L., Kemp, K., & Greenleaf, G. (2020). COVIDSafe App—Submission to the Parliamentary Joint Committee on Human Rights. Available at SSRN 3595109

Boughey, J., & Weeks, G. (2018). Government accountability as a "constitutional value." In R. Dixon (Ed.), Australian constitutional values (pp. 99-120). Hart Publishing.

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: Why care.data ran into trouble. Journal of Medical Ethics, 41(5), 404-409. https://doi.org/10.1136/medethics-2014-102374

Commission Nationale Informatique & Libertés (2017). How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence.

Commonwealth Ombudsman (2017). Centrelink's automated debt raising and recovery system: A report about the Department of Human Services' online compliance intervention system for debt raising and recovery (Investigation Report 2). https://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf

Council of Europe, Commissioner for Human Rights (2019). Unboxing Artificial Intelligence: 10 steps to protect Human Rights https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

Crawford, K., Whittaker, M., Elish, M. C., Barocas, S., Plasek, A., & Ferryman, K. (2016). The AI now report: The social and economic implications of artificial intelligence technologies in the near-term. https://ainowinstitute.org/AI_Now_2016_Report.pdf

Engineering and Physical Sciences Research Council (2010). Principles of robotics. https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press. https://books.google.com.au/books?id=pn4pDwAAQBAJ

Fjeld, J. (2020). Principled artificial intelligence: A map of ethical and rights-based approaches. Berkman Klein Center. https://cyber.harvard.edu/publication/2020/principled-ai

Gallie, W. B. (1964). Essentially contested concepts. In W. B. Gallie (Ed.), Philosophy and historical understanding (pp. 157-91). London: Chatto & Windus.

Guihot, M., & Moses, L. B. (2020). Artificial intelligence, robots, and the law. LexisNexis.

Henman, P. (2020). Improving public services using artificial intelligence: Possibilities, pitfalls, governance. Asia Pacific Journal of Public Administration [Ya Tai Gong Gong Xing Zheng Xue], 42(4), 209-21. https://doi.org/10.1080/23276665.2020.1816188

IEEE (n.d). The IEEE Global Initiative on ethics of autonomous and intelligent systems. https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014). A world that counts: Mobilising the data revolution for sustainable development. United Nations.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389.

Leins, K., Culnane, C., & Rubinstein, B. I. (2020). Tracking, tracing, trust: Contemplating mitigating the impact of COVID-19 through technological interventions. Medical Journal of Australia, 213(1), 6-8.e1. https://doi.org/10.5694/mja2.50669

Lupton, D. (2019). 'I'd like to think you could trust the government, but I don't really think we can': Australian women's attitudes to and experiences of My Health Record. Digital Health, 5, 2055207619847017. https://doi.org/10.1177/2055207619847017

Macnaghten, P., Davies, S. R., & Kearnes, M. (3 September, 2019). Understanding public responses to emerging technologies: A narrative approach. Journal of Environmental Policy & Planning, 21(5), 504-518. https://doi.org/10.1080/1523908X.2015.1053110

Mashaw, J. L. (2006). Accountability and institutional design: Some thoughts on the grammar of governance. In M. W. Dowdle (Ed.), Public accountability: Designs, dilemmas and experiences (pp. 115-56). Cambridge Studies in Law and Society. Cambridge UK: Cambridge University Press.

Mulgan, R. (2003). Holding power to account: Accountability in modern democracies. Basingstoke: Palgrave Macmillan.

New Zealand Government and Stats NZ (2020). Algorithm charter for Aotearoa New Zealand. https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

Newman, C., MacGibbon, J., Smith, A. K. J., Broady, T., Lupton, D., Davis, M., Bear, B., Bath, N., Comensoli, D., Cook, T., Duck-Chong, E., Ellard, J., Kim, J., Rule, J., & Holt, M. (2020). Understanding trust in digital health among communities affected by BBVs and STIs in Australia. https://doi.org/10.26190/5f6d72f17d2b5

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

OECD (2019). OECD principles on AI. https://www.oecd.org/going-digital/ai/principles/

O'Neill, O. (2002). A question of trust. Cambridge: Cambridge University Press.

Pang, P. C.-I., McKay, D., Chang, S., Chen, Q., Zhang, X., & Cui, L. (2020). Privacy concerns of the Australian My Health Record: Implications for other large-scale opt-out personal health records. Information Processing & Management, 57(6), 102364. https://doi.org/https://doi.org/10.1016/j.ipm.2020.102364

Rieder, G., & Simon, J. (2016). Datatrust: Or, the political quest for numerical evidence and the epistemologies of big data. Big Data & Society, 3(1), 2053951716649398. https://doi.org/10.1177/2053951716649398

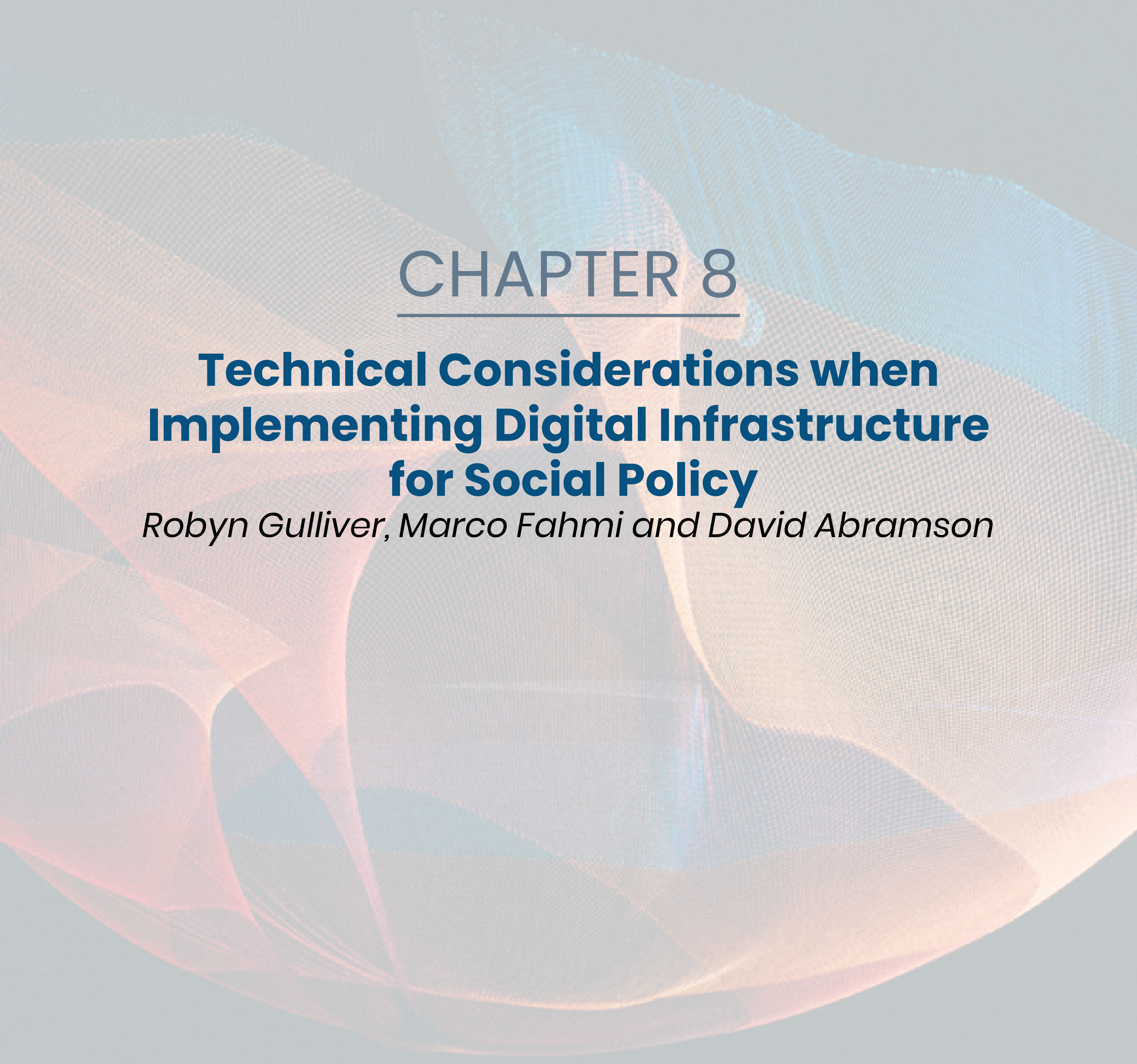
Sadowski, J. (2020). Too smart: How digital capitalism is extracting data, controlling our lives, and taking over the world. MIT Press.

Victoria Legal Aid (23 August 2019). Submission to the Senate Legal and Constitutional Affairs References Committee on the Inquiry into the Impact of Changes to Service Delivery Models on the Administration and Running of Government Programs. Melbourne: Victoria Legal Aid.

Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco—Or was it something worse? Australian Journal of Public Administration, 80(2), 340-360. https://doi.org/10.1111/1467-8500.12479

Zalnieriute, M., Bennett Moses, L., & Williams, G. (2019). The rule of law and automation of government decision-making. The Modern Law Review, 82(3): 425-455. https://doi.org/10.1111/1468-2230.12412

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile. https://books.google.com.au/books?id=W7ZEDgAAQBAJ



CHAPTER 8

Technical Considerations when Implementing Digital Infrastructure for Social Policy

Robyn Gulliver, Marco Fahmi and David Abramson

Technical Considerations when Implementing Digital Infrastructurefor Social Policy

INTRODUCTION

The incorporation of big data and advanced computational methods into social policy development and research is becoming increasingly common as we continue to experience rapid technological change (Carney, 2019). Yet despite the substantial body of research emerging on the characteristics and infrastructure challenges related to big data, integrating its use into social policy projects and research continues to present challenges to practitioners. As a result, there is an urgent need for identifying the technical considerations related to social policy and research. A better understanding of these technical infrastructure considerations will then enable policymakers and social sciences researchers to access, manipulate and operationalise their data securely and ethically.

While much research focuses on the challenges associated with the modelling and analysis phase (Agrawal et al., 2011), technical infrastructure choices can influence policy and research outcomes. Misuse of data and algorithmically-driven discrimination are just some of the potential consequences of ineffective digital infrastructure management when incorporating big data into social policy and research (Redden, 2017). To avoid these outcomes, technical infrastructure must offer a secure data storage environment, alongside integration of a range of analytical tools and datasets from multiple sources, while also complying with legislative and ethical requirements (Sivarajah et al., 2017).

In this chapter, we present a model designed to guide social policymakers and researchers through the technical considerations for integrating big data into their projects. This model identifies the technical considerations associated with six stages of the policy and research development cycle, namely, storage, integration, access, analysis, interpretation and operationalisation. We then demonstrate how the model can be used by presenting two example applications—the US E-Verify work permit system and Australia's COVIDSafe app. Through the development of this model, we aim to provide informative and actionable technical guidance for social policymakers and researchers seeking to use big data in their policy or research projects. Given the growing use of data in social policy and research alongside its potential for harm, a better understanding of how big data can be securely and safely accessed, analysed and operationalised is of significant value to the sector.

The use of big data in social policy and research

While the term "big data" has uncertain origins (Gandomi & Haider, 2015), scholars have identified six common characteristics: it is data which is high in volume, variety, velocity, veracity, value and variability. This data is composed of the digital by-products of human activity, composed of information gathered through activities such as digital transactions, social media use, sensor data and government functions (Coulton et al,, 2015). Big data has supported growth and competitive advantage in the business sector, and can also deliver substantial value in various social policy arenas (Günther et al., 2017). For example, effective integration of big data into health sector processes has been shown to enable patient-centric services, early detection of spread of disease, monitoring hospital quality and improving treatment methods (Archenaa & Anita, 2015), as well as the establishment of e-health integrated platforms (Black et al., 2014). Big data is also being increasingly integrated into policy making in diverse areas such as the prevention of child sexual abuse (Broadley, 2018) and the design of smart cities (Hong et al., 2019).

The escalating use of big data in policy and research has driven the development of a number of big data frameworks. In Australia, the federal government has implemented the Five Safes Framework, developed for the Office for National Statistics in the United Kingdom in 2002 (Commonwealth of Australia, 2019). This was designed to assist data custodians in balancing the risks and utility of using big data, and does so through highlighting five key areas for managing this risk (ABS, 2017) (see Table 8.1).

Table 8.1: Five Safes framework details

Safe	Key Question
Safe people	Is the researcher appropriately authorised to access and use the data?
Safe projects	Is the data to be used for an appropriate purpose?
Safe settings	Does the access environment prevent unauthorised use?
Safe data	Has appropriate and sufficient protection been applied to the data?
Safe output	Are the statistical results non-disclosive?

A number of frameworks and guides have also been developed in other countries and contexts to facilitate effective use of big data across different sectors. These include the Big Data Value Chain, which describes the information flow within any big data system (Curry, 2016), as well as sector-specific models (e.g., see Wang et al., 2018, for a detailed outline of a "big data analytics-enabled transformation model" applied to healthcare).

In addition to these models and frameworks, a growing number of open data portals such as the Secure eResearch platform provide users with secure environments to both access data and undertake data analytics (see Jones et al., 2019). These safe havens can maintain, verify and secure data, enabling new collaborations to generate insights from ethically acquired and used data.

Technical challenges of using big data in social policy and research

There are significant challenges associated with incorporating big data into social policy and research. The exponential growth in big data in such a short period of time means there is limited research on how big data can actually improve government service (Hong et al., 2019). In particular, how to weigh the potential benefits of big data in social policy and research against the significant technical and organisational resources required to safely and securely integrate it remains an open question (Carney, 2019).

While the use of big data requires careful infrastructure design, many public organisations may have limited funds, technical knowledge or evidence to help inform and develop comprehensive digital infrastructure (Daniell et al., 2016; Oussous et al., 2018). An organisation's legacy infrastructure may have resulted in the need to integrate multiple different technical systems across quite divergent departments, with multiple different compliance and application regimes (Coulton et al., 2015). Organisational structures or political priorities may interfere with the ability to effectively integrate big data into policy (Hong et al., 2019), as well as the ability to evaluate projects or incorporate new data to refine models and policies (Archenaa & Anita, 2015). Furthermore, a lack of technical protocols to manage data after use could result in inadvertent identification through data aggregation, or reuse of data in breach of data consent agreements.

Thus, there are considerable risks related to integrating big data into social policy safely. At the beginning of each particular project, there is a need for policymakers to start from first principles and probe what their existing infrastructure resources are, what technical requirements are needed to integrate big data into their project securely, and what benefits and risks will accrue. Many technical considerations will also be influenced by factors such as data governance arrangements, available resourcing and the personnel skills and abilities available to the organisation. While we focus on technical considerations in this model, we highlight where other considerations may also significantly impact on infrastructure choices. As a project is designed, consideration of the technical challenges of big data should be made at each of the multiple steps in the data analysis pipeline (Labrinidis & Jagadish, 2012). To assist with this work, we therefore present in the following sections a model with six components and associated considerations aligned with the data analysis pipeline.

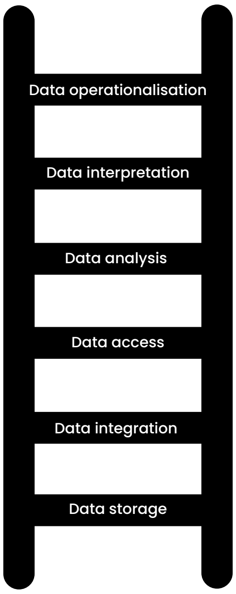
TECHNICAL CONSIDERATIONS MODEL

Researchers have presented a range of technical infrastructure designs for the use of big data in various contexts (Labrinidis & Jagadish, 2012). However, these solutions are often highly technical and specific to existing infrastructure capabilities. In practice, many social policymakers and researchers will be operating in diverse organisations with different infrastructure arrangements and technical constraints unique to their situation. Moreover, some social policymakers and researchers may not have any training in computational skills or technical design beyond standard statistics (Coulton et al., 2015).

Given this diversity, our model builds on a range of existing frameworks to identify technical considerations relevant for most policy and research projects, whether using legacy or new technical infrastructure. We do this through presenting technical considerations associated with each of the model's six components, followed by a set of questions for consideration. These questions can guide policymakers and researchers to interrogate their own technical infrastructure situation, and identify where risks may arise.

While these components build upon each other, this model can be used iteratively, with different components also able to be considered in isolation. That is, as new information arises (new data sources, techniques, technologies, applications, needs), the technical opportunities and constraints associated with each component may need to be revisited.

Figure 8.1. Technical considerations model



Component 1: Data storage

The first consideration of any social policy or research process involving big data is to specify what data is required, how it may be acquired, where it will be stored and how it will be integrated. While social policy and research has traditionally used data acquired from "structured" sources such as tax records, the vast majority of data available today is "unstructured" (e.g., see Cukier, 2010). Unstructured data is not necessarily organised into the traditional row and column structure of relational databases and can be overwhelming in size. Examples include social media data and smart meter sensor data (Labrinidis & Jagadish, 2012), which quickly overwhelm the storage capabilities of simple databases used in the past (Sivarajah et al., 2017). Projects where the use of unstructured big data is to be used must first ascertain what and how this specific data will be acquired, and whether existing digital infrastructure will be able to store and facilitate processing of increasing volumes of data and more complex datasets (Padgavankar & Gupta, 2014). A range of new storage options from commercial providers are now available which allow users to scale up storage quickly as required, and have virtually unlimited storage space. However, costs may be prohibitive, and technical skills in developing storage protocols and algorithms are still required (Yang et al., 2017).

Policymakers also need to consider how they wish to access the data. In general, there are two options for the logical organisation and transfer of data: a "data warehouse" or "federated data." A data warehouse is a centralised storage system where data from multiple administrative systems is brought together, linked, and stored into a centralised repository. Data warehouses can store vast volumes of heterogeneous and potentially sensitive data such as financial transactions, social media data and medical procedures, which need to be aggregated together (Sivarajah et al., 2017).

A federated model enables authorised users to access data which may be stored across multiple different systems and/or organisations. This decentralised architecture allows users to perform queries on the data within a federated network, often by application programming interfaces (APIs). Data results are returned to the user without ever leaving the organisation which stores it. In addition to these options, a range of hybrid models are also in use (see Yang et al., 2017) for a detailed overview of data warehouse architecture options).

The choice between centralised or federated systems depends on both the existing infrastructure arrangement as well as ethical and legal considerations. As such, many of the technical considerations relevant to this stage will be strongly influenced by existing data governance parameters within an organisation. Some projects may require integrated and stored data, and thus the centralised warehouse model is the only feasible option (Kumar, 2012). However, this model may incur significant organisational costs, given the need for specific technical skills and constant integration of growing unstructured data sources. Conversely, federated warehouses remove the need to store data, which may assist with managing legal issues associated with the use and storage of data, given that data remains in local control of the organisation that holds it (Herman, 2019).

New multiagency data portals are being developed specifically for policymakers and researchers to facilitate data storage and access. These portals offer pre-integrated data accessible under a range of different security parameters, and have significant potential for yielding insights able to transform social policy and practice (Coulton et al., 2015). They can function as a permanent data utility for the social sector and can deliver high-quality big data access systems (see Goerge et al., 2010; London & McLaughlin, 2014 for case studies). Policymakers and researchers could therefore also investigate whether commercial cloud systems or multiagency portals will meet the data needs for their projects.

The final consideration around data storage relates to data sharing agreements. There are strict regulations limiting data collection and use when dealing with highly sensitive data, such as in healthcare (Jagadish et al., 2014), or education (Cech et al., 2015). Data sharing agreements are necessary to clearly specify which data categories are to be shared, the conditions under which the data can be used, frequency of access, confidentiality and security requirements, and that these are adhered to (Kumar, 2012). Whether a federated, centralised or hybrid/cloud/portal system is used, the data manager needs to ensure data is stored and available to be used for policy and research purposes in a way consistent with their data-sharing agreement specifications.

Component 1 technical infrastructure considerations:

- Is the data we wish to use for our project structured or unstructured, what standards does it use, and how can it be acquired?
- What legacy storage infrastructure do we have and can it store this data?
- What regulations relate to the storage and use of this data?
- What parameters do our data sharing agreements/standards have for collecting and aggregating data with other sources we may want to use?

Component 2: Data integration

Component 2 relates to technical considerations associated with data integration. Data integration requires minimising redundant data, data consolidation and ensuring data consistency (Mosely et al., 2009). It can be a complex process, incorporating data cleaning and standardisation, along with checks for data quality and completeness measuring (for a detailed description, see Vatsalan et al., 2013). This process becomes particularly technically challenging if the project requires aggregating and integrating structured and unstructured data in large volumes into existing organisational infrastructures (Oussous et al., 2018).

Data integration can take place on a needs-based basis, periodically, or continuously (or, again, in hybrid forms of each). The choice of integration frequency will depend, firstly, on whether the organisation uses a data warehouse or federated data storage system (or hybrid), and secondly, on whether the project requires one-off data integration or ongoing data integration. Given that data integration is both complex and occurs throughout the big data system from beginning to end, effective integration is critical to maximising value from big data. Manuals such as the DAMA Guide to the Data Management Body of Knowledge (Mosely et al., 2009) provide in-depth information on how to plan and implement best-practice data management functions.

As with the storage considerations in Component 1, the choice of data integration processes involves choices and trade-offs. Federated and hybrid systems require sophisticated data integration protocols to obtain and match data acquired from different data depositories (Kumar, 2012). Furthermore, each data depository may have differing conditions of use imposed by a data sharing agreement or law. Even in data warehouse environments, integration will still have to occur between organisational units (e.g., client data and smart meter data), and as new data is acquired. In addition, efficient integration of large datasets may require automated pre-processing techniques and compression algorithms, and automatic generation of metadata to ensure transparent and secure data provenance information (Yang et al., 2017). Metadata can include the data's source of origin, format and details on retrieval and access (Blazquez & Domenech, 2018). Despite the development of numerous mathematical models for integrating records across data systems (e.g., see Kumar, 2012), developing and maintaining efficient and effective mining and cleaning of data remains challenging (Chen et al., 2013; Labrinidis & Jagadish, 2012).

Inadequate data integration can have severe consequences further downstream if data and metadata is low quality, if the volume of data exceeds infrastructure capacity, or if it is inconsistent and imprecise (Labrinidis & Jagadish, 2012). Poor-quality integration can significantly affect the capability of infrastructure systems to extract actionable information in the first place (Chen & Zhang, 2014), and lead to potentially incorrect policy operationalisation (Clarke, 2016). In addition, the process of data integration can reveal very personal and sensitive information, even when de-identified, raising questions of identity theft, illegal discrimination and unjust classification, as well as revealing personal and sensitive information which is at risk of public release (Günther et al., 2017). For example, Christensen and colleagues (2014) found in their analysis of Indigenous identification in data sources that a number of data quality issues, including missing data, inconsistency and error, could lead to errors in generating conclusions (see also Hunter & Ayyar, 2011). This was overcome by the development of a "Multi-Stage Median" algorithm, which resulted in less missing data and improved identification consistency. Thus, careful analysis and testing of data integration methods can facilitate improved data integration outcomes.

Component 2 technical infrastructure considerations:

- How often will data integration be required (e.g., on demand, continuously or periodically)?
- How will be the data be standardised/normalised and de-identified?
- Is there any missing/incorrect/skewed data and how can this be addressed and rectified?
- How can the data integration be tested to ensure completeness and accuracy?
- What could be the consequences of poor-quality data integration for our project?

Component 3: Data access

Once data is cleaned, matched and de-identified, technical considerations related to access then arise. There are numerous benefits from sharing and accessing data (Günther et al., 2017), as accessing information derived from the integration of big data is a key area of delivering valuable social policy insights (Coulton et al., 2015).

There are two main options for delivering integrated data: centralised gateways or decentralised points of access. Centralised gateways can exist internally in the organisation or in a cloud computing environment, and enable secure, controlled access in compliance with governance and management politics. The gateways may make data management and access easier (Miranda et al., 2015) and ensure that only those with prior permission can access data relevant for their policy or research project (Sivarajah et al., 2017).

In a decentralised access environment, organisations can use a range of mechanisms to ensure that data sharing between teams or external groups is controlled and that data sharing agreements are adhered to. These include ensuring security protocols are embedded in the data management system and maintaining a robust process for identifying and tracking which users access the data (Kennedy & Moss, 2015; van den Broek & van Veenstra, 2015). These mechanisms can reduce insecure access and leave a data trail for investigating unauthorised data access or use.

In addition to determining access rights and protocols, users need to be able to efficiently query and retrieve the data they require (Peek et al., 2014). Access systems should therefore enable users to view and search a catalogue of available datasets and should include data descriptions (metadata) and the publisher in order to verify source authenticity (Lněnička & Máchová, 2015). Enabling this is particularly challenging in decentralised information delivery systems, where policymakers and researchers may gain access to integrated data via different systems or through different organisational units. In this situation, more complex security protocols and secure authentication systems may be required. These will need to both protect the data and the security of the organisational network (Archenaa & Anita, 2015).

Recent research highlights the increasing adoption of cloud-based information delivery environments for a range of social policy uses (e.g., see Singh et al., 2020). This is partly because these environments may be more secure, offer more user friendly access for some types of data, and be quickly scalable as data storage and processing needs change. This is because these web-based systems collect data from multiple sources, which is then searchable and accessible via an API. They can provide a single point of access for government data (for a review, see Lnenicka, 2015). In Australia, a number of online secure labs such as the Australian Bureau of Statistics DataLab provide data access and analysis services for authorised users. Similarly, some cities have built open data portals utilising APIs which allow the combination of data from a variety of sources such as corporate foreclosures, evictions and deed transfers. This enables foreclosure prevention specialists to more efficiently access and integrate data sources to help fight predatory lending from a single point (Coulton et al., 2010; Nelson, 2014). As such, open data portals have shown promise in increasing accessibility to big data for developing social policy while preserving data security.

Component 3 technical infrastructure considerations:

- What information delivery systems already exist in our organisation?
- Who needs access to the data, and how will they find the data they need?
- What security protocols are needed to limit and track access?
- How will unauthorised access be identified and blocked?

Component 4: Data analysis

Data analysis involves the use of analysis tools such as natural language processing software, and analysis techniques such as inferential statistics and predictive modelling. A range of different analysis tools and techniques are employed to analyse large-scale, complex data for various applications, and this number is growing continuously. Two specific technical considerations relate to this issue. First, policymakers and researchers must ensure there is sufficient infrastructure capability to undertake analysis with high volumes of data. Second, they must ensure that their technical infrastructure is able to apply analytical algorithms which can cope with the diversity and complexity of big data. In addition, technical choices related to data analysis may also be strongly influenced by the skills and resources available to the organisation.

First, big data analysis may require combining queries across datasets (Peek et al., 2014), and, as such, dedicated computing resources with adequate CPU, network and storage capacity must be in place to enable data analytics to occur (Chen et al., 2013). The processing resources required for big data can far exceed the processing power offered by traditional organisational systems and the ability of basic analysis software to process. This is exacerbated by the fact that some analysis of large datasets needs to be undertaken in an automated manner, due to its very large volume. However, whether automated analysis is possible depends on the location where the data is hosted, the type of database it is hosted in, and the particular type of analysis being undertaken (Labrinidis & Jagadish, 2012). One solution to this issue is to use centralised cloud-based systems, which can provide sufficient storage and processing power while offering advanced analytical tools (Yang et al., 2017). Depending on data storage and sharing agreements, cloud-based systems may enable large datasets to be processed on remote internet servers rather than researchers' or organisations' own computers. Conversely, the choice of an internal centralised storage and access system may encourage organisations to fund positions dedicated to data analysis. For example, some organisations have created "competency centres" examples, where one particular organisational unit offers expertise in analytics and services (Günther et al., 2017). Centralised systems can also provide technical skills assistance where needed, and foster multidisciplinary work (Gao et al., 2015).

The second technical consideration associated with this component is ensuring that big datasets are processed and analysed in an appropriate way. Data analysis of big datasets requires working with highly variable and heterogeneous data. As a result, pre-processing of data into structured form for analysis may be required. A growing number of technical solutions, such as Apache's Hadoop big data processing platform, are enabling parallel processing platforms to overcome this step (Apache Software Foundation, 2010). Even so, the emerging use of artificial intelligence and deep learning systems may require very high processing capabilities, continuous access to data, and bulk storage in order to run analyses (Pritchard, 2019). As a result, some researchers argue that creating highly efficient and scalable data reduction algorithms to accelerate analysis of big data is one of the most important tasks in big data research (Zhai et al., 2014). Accordingly, a range of data reduction packages have been developed for different software platforms (Oussous, 2018) as well as for specific social policy uses (e.g., see Wang et al., 2018 regarding data reduction options for smart city big data). For social policymakers and researchers involved in one-off projects, or operating with little technical assistance, awareness of these technical considerations associated with data analysis can help identify and alleviate potential bias and errors before outputs are used to inform policy.

Component 4 technical infrastructure considerations:

- Does our organisation's existing digital infrastructure provide relevant data analysis tools and support relevant analysis techniques?
- Does our organisation have a data analysis centre or provide analysis expertise?
- If no expertise is available in the organisation, is there a cloud-based system offering the data analysis tools and techniques required?
- Does the data need to be pre-processed prior to undertaking data analysis?

Component 5: Data interpretation

Component 5 addresses the technical considerations related to using the data analysis outputs to create meaningful, equitable and ethical policy proposals. Given the importance of accurate data interpretation in the social policy and research context, and an increasing range of tools facilitating interpretation of large, complex big datasets, we treat "interpretation" as a standalone component in this technical infrastructure model. While a large body of research highlights the potential for biased data interpretations from big data analytic outputs (Turner & Resnick, 2019), technical infrastructure design can enable output testing and meaningful visualisation to reduce the likelihood of misinterpretation.

It is important that the technical infrastructure design supports testing of data analysis outputs. This can be done through establishing systems to undertake policy simulations based on analysis outputs (Daniell et al., 2016). Options include the incorporation of "sandbox" environments where simulations can be run to check for biased outputs, or test for unforeseen negative impacts prior to policy implementation. In addition, technical infrastructure should support new algorithmic tests for reducing bias, which are under continuous development. For example, the Reducing Bias Amplification is a framework imposed on algorithms which constrain the algorithm from generated biased outputs, such as where an AI recruitment program preferentially selected men for technical roles at Amazon (Williamson, 2019). Similar solutions such as the “Endor” analysis platform can analyse large quantities of data to run predictions while filtering out existing bias such as gender (Williamson, 2019). These additional analytical capabilities can reduce the likelihood that high-cost policies could lead to negative and unforeseen consequences by enabling testing, and can be designed into the technical infrastructure to connect with analysis process and outputs.

The second technical consideration related to this component is whether visualisation capabilities are embedded in the technical infrastructure. Visualisation capability can be important for interrogating data analysis outputs. This is particularly valuable, given the multi-dimensional nature of big data, where insights may be gleaned across factors which differ in location, time, and demographics. Visualisation can uncover hidden patterns and unknown correlations, and is critical to make sense of big data (Chen & Zhang, 2014; Padgavankar & Gupta, 2014). However, the outputs of data analysis can be distorted through incorrect visualisation and oversimplification of the analysis output (Olshannikova et al., 2015). Adequate visualisation requires key infrastructure components, namely: an interface to filter or sample data prior to visualisation, processing capability, simple and intuitive visual analytic tools and portals for sharing or viewing visualisations. Fulfilling these technical requirements requires fusing multiple data sources of heterogeneous data which may also have spatial components, which can be technically challenging (Reda et al., 2013). In addition, many visualisation programs (such as Gephi, ArcGIS or Tableau) will require the users to learn new software, which can be costly in terms of finance and time.

Component 5 technical infrastructure considerations:

- What tools exist within our digital infrastructure to test the validity of data analysis outputs?
- Are there any relevant bias reduction strategies which can be used during the data analysis phase?
- Are there tools available to run simulation tests or test the impact of analysis outputs in a secure sandbox?
- How can the data analysis outputs be visualised?
- Are the visualisations capturing the complexity of the analysis outputs or overly simplifying them?

Component 6: Data operationalisation

The final component of our model considers the technical challenges which arise through operationalisation of the policy or research outcomes. There are three primary infrastructure considerations at this layer: replication and transparency, post-analysis data management, and de-identification/incorporation of additional data.

The ability to independently replicate scientific findings is a fundamental requirement of research (Hastings et al., 2019). The systems and processes by which this can occur will depend on the technical infrastructure arrangements identified in Components 1 and 2 of the model, as well as the data access options considered in Component 3. If a centralised warehouse is used, then identifiable data may continue to be stored within databases (depending on data-sharing agreements) and thus be able to be used to replicate projects in the future. In this case, any potentially identifiable integrated data must be tagged via metadata to ensure it is not used for purposes other than that permitted. Under a federated system, project related data may have been integrated, analysed and stored on the analysts’ own computers. Thus, this data may need to be deleted or securely stored in a different digital environment, depending on data access and sharing agreements.

Given that each situation may differ, a process for preserving privacy and determining access rights according to data-sharing agreements should be built into the digital infrastructure environment itself. This process would ensure transparency, project replication, and accessing of data where applicable for future projects. One example of an effective system was presented by Hastings and colleagues (2019), who developed a system for post-operationalisation record preservation, designed to preserve privacy while enabling replication. Their system involved preserving de-identified data, which was linked to research/project versions and then validated by the project team. Once validated, the identifiable data and integrated data was wiped, with any further analysis (as required) done on the de-identified data.

Those who are most affected by policy outputs generated through big data analysis may find accessing project information complex and opaque (Höchtl et al., 2016). While transparency outside the organisation may be restricted by necessity, standardised templates for code and data manipulation should be tied to projects to enable replication within the organisation or by other researchers. This will also enable incorporation of new data into the project in the future. A significant benefit of the growth of big data is the ability to incorporate both new data, and new automated methods, tools and techniques into the evaluation and refinement process (Höchtl et al., 2016). This is not only to enable incorporation of additional data into the policy cycle, but also to allow for data which can quickly become superseded (e.g., census data). As Höchtl and colleagues (2016) suggest, effective infrastructure allowing data integration from different datasets could allow data updating and project assessments to be undertaken much more frequently. This then enables continuous evaluation to be part of the policy cycle, as well as enabling post-project implementation options such as tracking outcomes and monitoring compliance. Finally, consideration of whether project findings may result in individuals being re-identified is critical both for protecting the privacy of the individual and for reducing the risk of disclosing stigmatising group attributes. A range of guides have been developed for designing de-identification into the decision-making framework (e.g., see O'Keefe et al., 2017), thus ensuring these processes are incorporated into the technical framework will help minimise these risk.

Finally, the execution of new policies will produce new data, which can then help evaluate policy effectiveness and identify problems at an early stage of implementation. Ensuring that the technical infrastructure enables incorporation of new data such as a social media posts, economic data or policing records can not only help improve policy, but also can help ascertain public opinion on changes and unexpected responses (Alfaro et al., 2016). Transparent and secure records of the data analysis process can thus enable rerunning projects as required, enabling the continuous improvement and evolution of social policy and research outcomes.

Component 6 technical infrastructure considerations:

- What is the process for managing data once the project has been implemented?
- Is this process embedded in the digital infrastructure, ensuring that data sharing and access rights are respected post-project?
- What documentation is needed to ensure project analysis transparency and replicability?
- What data could be captured to evaluate post-implementation?
- Can this data be integrated into the existing data storage and integration system?

APPLICATIONS

In this section, we apply the model to two examples relevant to social policy issues: the E-Verify employment database in the US and the Australian COVIDSafe app. These were selected due to both their heavy reliance on technical infrastructure, and the significant impacts technical considerations have had on those who use the systems. In the following sections, we present an overview of both E-Verify and the COVIDSafe app. We then consider the key technical infrastructure concerns related to each of the six components of our model and the consequences of those choices post-implementation.

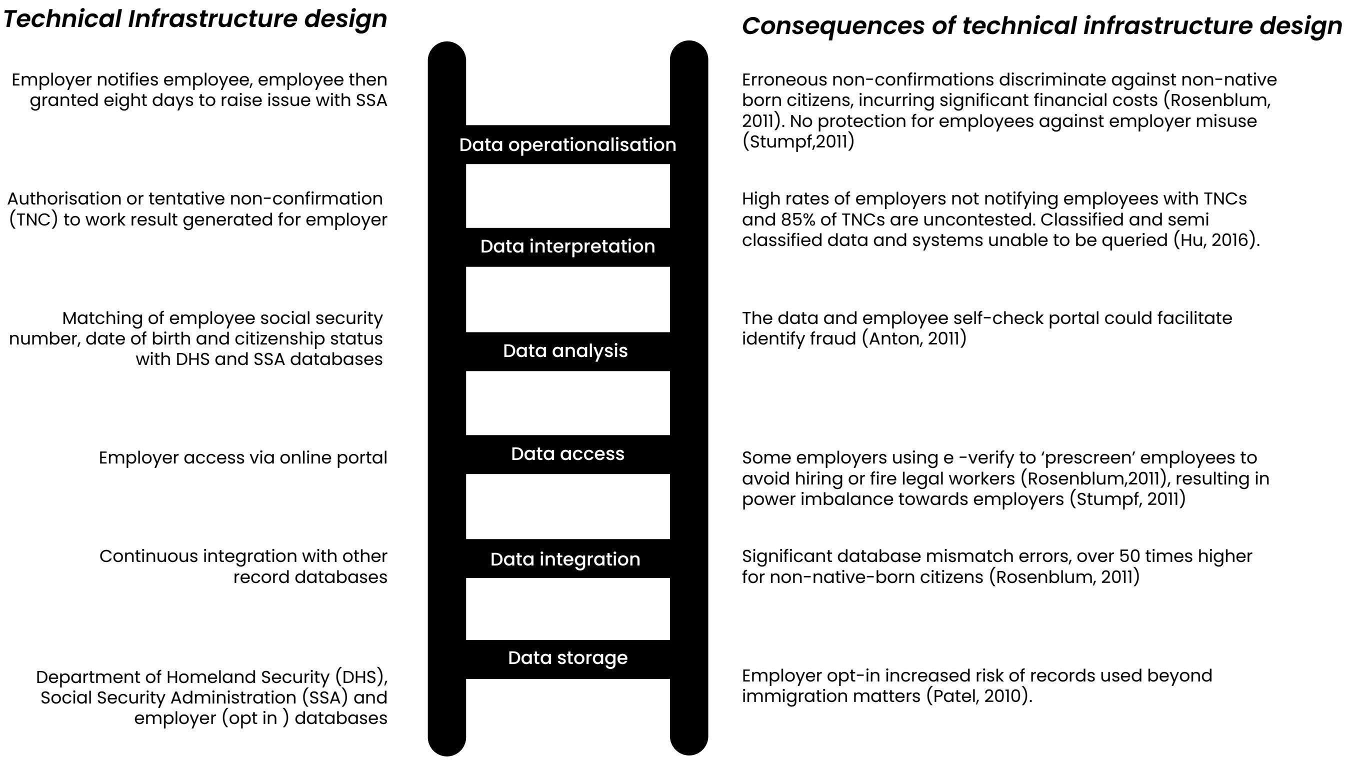
The examples are presented to demonstrate how the model can be used to identify potential issues and understand how technical considerations can lead to particular policy-related consequences. As they highlight, data infrastructure for social policy links social and technical components, and these components strongly influence each other when it comes to how the infrastructure is designed and used.

These examples also demonstrate how our model can be applied at the planning stage to identify issues before they arise, as well as retrospectively to analyse failings after project implementation. Ongoing reviews allow identification of ongoing technical infrastructure improvements related to each component, as demonstrated in both examples. For example, both E-Verify and COVIDSafe have undergone reviews of their infrastructure design. These have resulted in improved infrastructure arrangements related to the first two components of our model of data storage and integration (e.g., by enabling the addition of more pre-digitalised documents in the E-Verify system) as well as considerations related to Component 3, addressing data access concerns such as secondary use of the data (e.g., the COVIDSafe privacy impact assessment process). Finally, these examples highlight how the effectiveness of technical infrastructure choices can be affected by the real-world context. For example, while the choice of continuous data integration for E-Verify records allowed unprecedented data integration, the difficulty of ensuring paper records were accurate and matchable led to substantial mismatches, with significant negative effects.

E-Verify

E-Verify is an internet-based database developed to enable employers to ascertain employment rights of prospective employees through an online portal. The system was designed to increase employment process efficiency as well as reduce appearance-based discrimination against prospective employees (Stumpf, 2011). It combines data across multiple datasets into the Department of Homeland Security (DHS) and Social Security Administration (SSA) systems and was used by over 238,000 employers by 2011 (Antón, 2011). Various commentators have highlighted the three technical infrastructure requirements of E-Verify. It needed to ensure accurate identification and verification of an individual’s work rights (Antón, 2011). It needed to ensure that data could not be used by other agencies or be misused (Patel, 2010), and finally, it needed transparent processes in order to identify erroneous outcomes and then rectify any problems which arise (Hu, 2016). Figure 8.2 presents the technical infrastructure design alongside the consequences of those design choices, observed after E-Verify was rolled out using our model.

Figure 8.2. E-Verify system technical infrastructure design and consequence



Technical Infrastructure design		*Consequences of technical infrastructure design*
Employer notifies employee, employee then granted eight days to raise issue with SSA	Data operationalisation	Erroneous non-confirmations discriminate against non-native born citizens, incurring significant financial costs (Rosenblum, 2011). No protection for employees against employer misuse (Stumpf,2011)
Authorisation or tentative non-confirmation (TNC) to work result generated for employer	Data interpretation	High rates of employers not notifying employees with TNCs and 85% of TNCs are uncontested. Classified and semi classified data and systems unable to be queried (Hu, 2016).
Matching of employee social security number, date of birth and citizenship status with DHS and SSA databases	Data analysis	The data and employee self-check portal could facilitate identify fraud (Anton, 2011)
Employer access via online portal	Data access	Some employers using e -verify to ‘prescreen’ employees to avoid hiring or fire legal workers (Rosenblum,2011), resulting in power imbalance towards employers (Stumpf, 2011)
Continuous integration with other record databases	Data integration	Significant database mismatch errors, over 50 times higher for non-native-born citizens (Rosenblum, 2011)
Department of Homeland Security (DHS), Social Security Administration (SSA) and employer (opt in) databases	Data storage	Employer opt-in increased risk of records used beyond immigration matters (Patel, 2010).

As highlighted in Figure 8.2 above, the choices made regarding the technical infrastructure supporting the E-Verify system have led to a number of detrimental consequences. The most significant of these is the rate of database mismatches, with 4.1% of records (17.8 million in total) showing errors in name, citizenship status or date of birth (Hu, 2016). Mismatches resulting from faulty data-matching processes were shown in one study to occur 98 times more frequently for naturalised citizens than native-born US citizens (Hu, 2016). As a result, up to 189,000 US citizens and authorised users were denied employment opportunities as a result of erroneous E-Verify decisions (Hu, 2016). As shown in Figure 8.2, the negative outcomes of this system are compounded by other technical infrastructure choices, giving employers greater power over employees through their access to the portal, and the opaque nature of DHS and SSA databases and matching algorithms (Stumpf, 2011).

While some researchers have argued that the E-Verify system has at least partially achieved its goals of increasing efficiency and reducing discrimination (Rosenblum, 2011), design and implementation of technical solutions remains a high priority (Hu, 2016). Increasing awareness of these inequitable outcomes has occurred, and various proposals and projects have been undertaken to improve the infrastructure design and process. These include the incorporation of more pre-digitalised data such as passport data and photographs (Anton, 2011), requirements that employers double check for data entry errors (Rosenblum, 2011) and proposals to use more secure authentication options than the Social Security number (Anton, 2011).

However, each of these options involves balancing increased accuracy and individual protections against cost and infrastructure capabilities. Given that a mandatory electronic verification system could cost over $6 billion (Rosenblum, 2011) while potentially magnifying the negative outcomes experienced by its users, resolving these technical issues is of high priority. By using our model, choices around infrastructure components can be investigated as to whether they will generate outcomes which may undermine the secure and ethical use of the E-Verify system. While technical issues related to each component could have been quickly identified and developed prior to its release, the model can also be used on an ongoing basis to assess future proposals for improving E-Verify in the future.

Australia's COVID-19 app

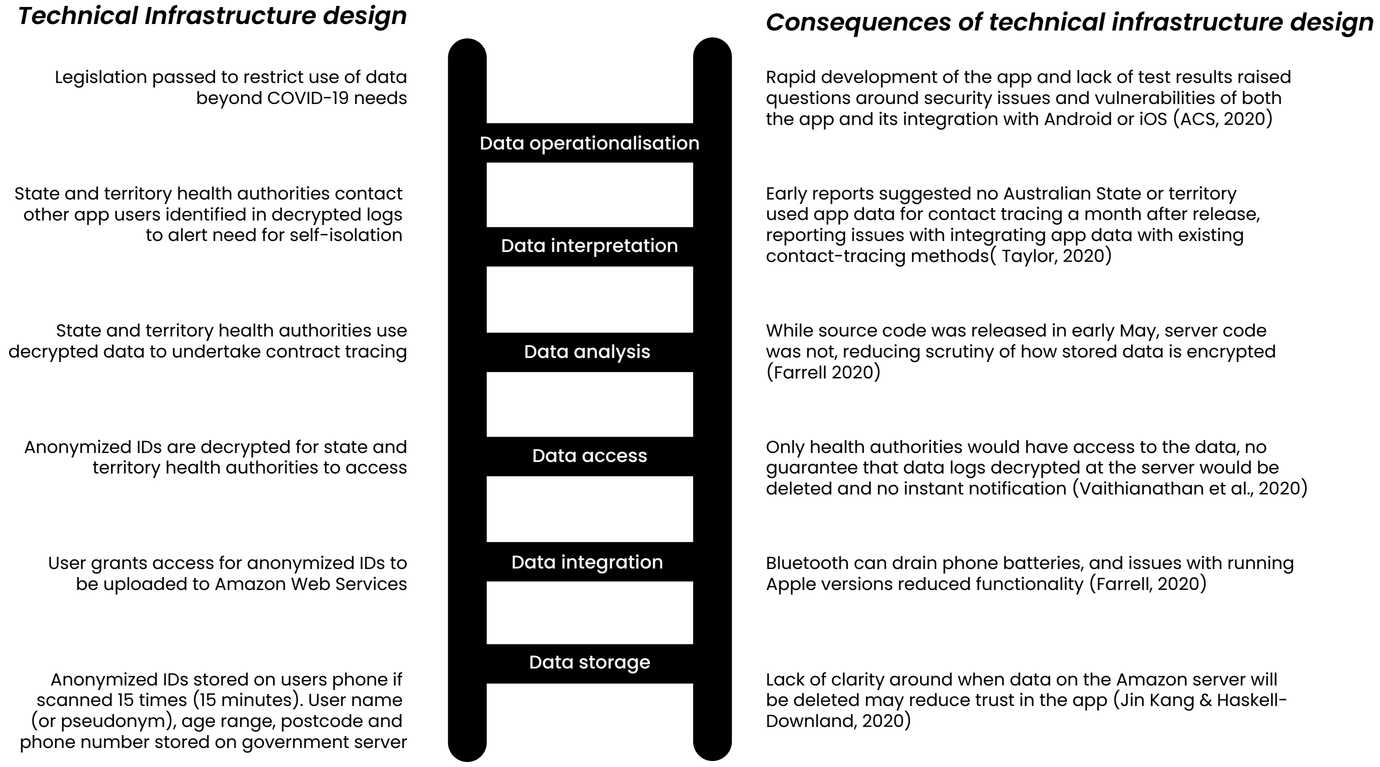
With the declaration of the COVID-19 pandemic in early 2020 by the World Health Organisation, many governments around the globe began rapid development of a track and trace digital application tool. The Australian Government developed COVIDSafe by building on the Singapore TraceTogether app. This app uses Bluetooth to collect anonymised ID records of other people with whom the individual has been within 1.5m for more than 15 minutes. This data, stored on the user's phone for 21 days, is then uploaded (after consent is granted by the user) to a commercial cloud service (the Amazon Web Services server in Australia) for state health authorities to utilise for contract tracing when an infection is detected.

The app had one overarching goal: to facilitate more effective track and trace processes to contain the pandemic. Figure 8.3 presents the technical infrastructure of the COVIDSafe app and consequences resulting from these infrastructure choices aligned with the components of the model.

Since its implementation, a number of researchers and experts have undertaken reviews of the app (e.g., see Pearcey Foundation, 2020) and highlighted successful infrastructure choices which protect privacy (such as regularly anonymised IDs) and maintain users' control (such as permission requirements for uploading data to the cloud service). However, as highlighted in our model in Figure 8.3, the rapid development and deployment of the COVIDSafe app led to questions regarding whether enough testing had been undertaken to identify security issues (ACS, 2020) and why the survey code was not released (Farrell, 2020). In addition, problems relating to impaired functionality arose (particularly on the Apple version), and some researchers identified a lack of clarity around when data uploaded to the server would be deleted (Jin Kang & Haskell-Dowland, 2020). Given that public acceptance of the app was critical to ensure enough individuals choose to download and utilise it, these issues may have resulted in reduced public trust in the app.

Some of these infrastructure considerations may be linked back to the short time frame required to design and implement the app during the pandemic crisis. However, even under substantial time constraints, the model could have assisted in identifying negative outcomes of certain infrastructure choices and point to solutions which prioritise public trust. These solutions could then have been efficiently identified and implemented. For example, our model indicates that the timely release of source code and transparent information sharing around the use of downloaded data could have helped strengthen public trust in the app. Our model therefore concisely categorises which technical infrastructure solutions could be prioritised, and thus help ensure the app fulfils its purpose, enabling an efficient and effective response to the COVID-19 pandemic.

Figure 8.3. Australian COVIDSafe app technical infrastructure design and consequences



Technical Infrastructure design		*Consequences of technical infrastructure design*
Legislation passed to restrict use of data beyond COVID-19 needs	Data operationalisation	Rapid development of the app and lack of test results raised questions around security issues and vulnerabilities of both the app and its integration with Android or iOS (ACS, 2020)
State and territory health authorities contact other app users identified in decrypted logs to alert need for self-isolation	Data interpretation	Early reports suggested no Australian State or territory used app data for contact tracing a month after release, reporting issues with integrating app data with existing contact-tracing methods(Taylor, 2020)
State and territory health authorities use decrypted data to undertake contract tracing	Data analysis	While source code was released in early May, server code was not, reducing scrutiny of how stored data is encrypted (Farrell 2020)
Anonymized IDs are decrypted for state and territory health authorities to access	Data access	Only health authorities would have access to the data, no guarantee that data logs decrypted at the server would be deleted and no instant notification (Vaithianathan et al., 2020)
User grants access for anonymized IDs to be uploaded to Amazon Web Services	Data integration	Bluetooth can drain phone batteries, and issues with running Apple versions reduced functionality (Farrell, 2020)
Anonymized IDs stored on users phone if scanned 15 times (15 minutes). User name (or pseudonym), age range, postcode and phone number stored on government server	Data storage	Lack of clarity around when data on the Amazon server will be deleted may reduce trust in the app (Jin Kang & Haskell-Downland, 2020)

CONCLUSION

Big data offers unprecedented opportunities to create new social policy tools to deliver outcomes which can improve countless lives. However, alongside this potential for good are significant risks for causing harm. Accordingly, some scholars have argued that the most significant challenge of using big data for social policy is not any lack of data, but instead, how to identify and implement big data solutions which support decision making (Archenaa & Anita, 2015). The success of these solutions depends on appropriate digital infrastructure design (Blazquez & Domenech, 2018).

In this chapter, we have presented a model to assist social policymakers and researchers in ascertaining the most effective digital infrastructure design choices specific for their projects. This model, composed of six components, highlights key technical considerations associated with each stage of the data analysis pipeline, followed by a set of questions for consideration by project designers. Through presenting this model, we hope to assist policymakers in balancing the risks and benefits associated with different technical infrastructure choices. This will help inform more successful policy and research outputs. In addition, increased understanding of the technical considerations related to big data will support more effective collaboration between policymakers and researchers, as well as increase data-driven innovation.

ACKNOWLEDGEMENTS

We would like to thank Janet Chan, Peter Saunders, Janeen Baxter, Ross Homel, Sheila Shaver and Mark Western for their work as co-investigators on this Linkage Learned Academies Special Project. This chapter has been published by the Australian Journal of Social Issues (volume 56, number 2). It is reproduced here with the permission of Wiley under licence 5121680074608.

REFERENCES

ABS (2017, August). 1160.0 - ABS Confidentiality Series. https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017

ACS (2020). TraceTogether mobile application. https://www.acs.org.au/search.html?q=covidsafe

Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., . . . Han, J. (2011). Challenges and opportunities with big data 2011-1. Cyber center technical reports. https://docs.lib.purdue.edu/cctech/1/

Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. Annals of Operations Research, 236(1), 197-213.

Antón, A. (2011). Social security administration’s role in verifying employment eligibility. CSC Initiative, Compliance Org.

Apache Software Foundation (2010). Hadoop. https://hadoop.apache.org

Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. Procedia Computer Science, 50, 408-413.

Black, A. S., Sahama, T. R., & Gajanayake, R. (2014). ehealth-as-a-service (ehaas): A data-driven decision making approach in Australian context. Studies in Health Technology and Informatics [e-Health-For Continuity of Care], 205, 915-919.

Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. Technological Forecasting and Social Change, 130, 99-113.

Broadley, K. (2018). What can surveillance data and risk factor research contribute to a public health approach to preventing child sexual abuse? Australian Journal of Social Issues, 53(4), 372-385. https://doi.org/10.1002/ajs4.46

Carney, T. (2019). Automation in social security: Implications for merits review? Australian Journal of Social Issues, 1-15. https://doi.org/10.1002/ajs4.95

Cech, T., Spaulding, T., & Cazier, J. (2015, August). Applying business analytic methods to improve organizational performance in the public school system. Paper presented at the Proceedings of the Twenty-First Americas Conference on Information Systems, Puerto Rico.

Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences, 275, 314-347.

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: A data management perspective. Frontiers of Computer Science, 7(2), 157-164.

Christensen, D., Davis, G., Draper, G., Mitrou, F., McKeown, S., Lawrence, D., . . . Zubrick, S. R. (2014). Evidence for the use of an algorithm in resolving inconsistent and missing Indigenous status in administrative data collections. Australian Journal of Social Issues, 49(4), 423-443. https://doi.org/10.1002/j.1839-4655.2014.tb00322.x

Clarke, R. (2016). Big data, big risks. Information Systems Journal, 26(1), 77-90.

Commonwealth of Australia. (2019). Best practice guide to applying data sharing principles. Commonwealth of Australia. https://www.pmc.gov.au/sites/default/files/publications/data-sharing-principles-best-practice-guide-15-mar-2019_0.pdf

Coulton, C., Goerge, R., Putnam-Hornstein, E., & de Haan, B. (2015). Harnessing big data for social good: A grand challenge for social work, Working Paper No. 11. Cleveland: American Academy of Social Work and Social Welfare.

Coulton, C., Schramm, M., & Hirsh, A. (2010). REO and beyond: The aftermath of the foreclosure crisis in Cuyahoga County, Ohio. Case Western Reserve University.

Cukier, D. (2010). Data, data everywhere: A special report on managing data. https://www.economist.com/special-report/2010/02/27/data-data-everywhere

Curry, E. (2016). The big data value chain: Definitions, concepts, and theoretical approaches. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), New horizons for a data-driven economy (pp. 29-37). Springer, Cham.

Daniell, K. A., Morton, A., & Insua, D. R. (2016). Policy analysis and policy analytics. Annals of Operations Research, 236(1), 1-13.

Farrell, P. (2020, 14 May). Experts raise concerns about security of coronavirus tracing app COVIDSafe. ABC News. https://www.abc.net.au/news/2020-05-14/experts-concerned-about-coronavirus-tracing-covidsafe-security/12245122

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144.

Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in big data analytics projects. Twenty-first Americas Conference on Information Systems, Puerto Rico.

Goerge, R. M., Smithgall, C., Seshadri, R., & Ballard, P. (2010). Illinois families and their use of multiple service systems. Chicago: Chapin Hall at the University of Chicago.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems, 26(3), 191-209.

Hastings, J., Michie, S., & Johnston, M. (2019). Theory and ontology in building cumulative behavioural science. https://osf.io/9te3x/

Herman, A. (2019). Federated data systems: Balancing innovation and trust in the use of sensitive data, White Paper. World Economic Forum.

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. Journal of Organizational Computing and Electronic Commerce, 26(1-2), 147-169.

Hong, S., Hyoung Kim, S., Kim, Y., & Park, J. (2019). Big data and government: Evidence of the role of big data for smart cities. Big Data & Society, 6(1), 2053951719842543.

Hu, M. (2016). Big data blacklisting. Florida Law Review, 67, 1735.

Hunter, B., & Ayyar, A. (2011). Undercounts in offender data and closing the gap between Indigenous and other Australians. Australian Journal of Social Issues, 46(1), 69-89.

Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. Communications of the ACM, 57(7), 86-94.

Jin Kang, J., & Haskell-Dowland, P. (2020). How safe is COVIDSafe? What you should know about the app's issues, and Bluetooth-related risks. The Conversation. https://theconversation.com/how-safe-is-covidsafe-what-you-should-know-about-the-apps-issues-and-bluetooth-related-risks-137894

Jones, K. H., Ford, D. V., Thompson, S., & Lyons, R. (2019). A profile of the SAIL Databank on the UK Secure Research Platform. International Journal of Population Data Science, 4(2).

Kennedy, H., & Moss, G. (2015). Known or knowing publics? Social media data mining and the question of public agency. Big Data & Society, 2(2), 2053951715611145.

Kumar, P. (2012). An overview of architectures and techniques for integrated data systems (IDS) implementation. Integrating Factors Inc.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. Proceedings of the VLDB Endowment, 5(12), 2032-2033.

Lee, N. T., Resnick, P.,& Barton, G. (2019). Algorithms bias detection and mitigation: Best practices and policies to reduce consumer harms. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Lnenicka, M. (2015). An in-depth analysis of open data portals as an emerging public e-service. International Journal of Social, Education, Economics and Management Engineering, 9(2), 589-599.

Lněnička, M., & Máchová, R. (2015). Open (big) data and the importance of data catalogs and portals for the public sector. Paper presented at the 3rd International Global Virtual Conference.

London, R. A., & McLaughlin, M. (2014). The youth sector: Supporting cross-institutional community collaboration through shared data. In N. Cytron, K. L. S. Pettit, & G T. Kingsley, What counts: Harnessing data for America's communities (pp. 177-189). Federal Reserve Bank of San Francisco & the Urban Institute.

Miranda, S., Kim, I., & Wang, D. D. (2015). Whose talk is walked? IT decentralizability, vendor versus adopter discourse, and the diffusion of social media versus big data. Missouri University of Science and Technology. https://scholarsmine.mst.edu/bio_inftec_facwork/389/

Mosely, M., Brackett, M., Early, S., & Henderson, D. (2009). The data management body of knowledge (DAMA-DMBOK guide). The Data Management Association, USA.

Nelson, S. C. (2014). Playing favorites: How shared beliefs shape the IMF's lending decisions. International Organization, 68(2), 297-328.

O'Keefe, C. M., Otarepec, S., Elliot, M., Mackey, E., & O'Hara, K. (2017). The de-identification decision-making framework. CSIRO Reports EP173122 and EP175702. https://www.data61.csiro.au/en/Our-Work/Safety-and-Security/Privacy-Preservation/De-identification-Decision-Making-Framework

Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing big data with augmented and virtual reality: Challenges and research agenda. Journal of Big Data, 2(1), 22.

Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., & Belfkih, S. (2018). Big data technologies: A survey. Journal of King Saud University-Computer and Information Sciences, 30(4), 431-448.

Padgavankar, M., & Gupta, S. (2014). Big data storage and challenges. International Journal of Computer Science and Information Technologies, 5(2), 2218-2223.

Patel, S. C. (2010). E-Verify: An exceptionalist system embedded in the immigration reform battle between federal and state governments. BC Third World LJ, 30, 453.

Pearcey Foundation (2020). The COVIDSafe App: A question of trust. https://pearcey.org.au/covidsafe-app-paper/

Peek, N., Holmes, J. H., & Sun, J. (2014). Technical challenges for big data in biomedicine and health: Data sources, infrastructure, and analytics. Yearbook of Medical Informatics, 23(01), 42-47.

Pritchard, S. (2019). AI storage: Machine learning, deep learning and storage needs. https://www.computerweekly.com/feature/AI-storage-Machine-learning-deep-learning-and-storage-needs

Reda, K., Febretti, A., Knoll, A., Aurisano, J., Leigh, J., Johnson, A., . . . Hereld, M. (2013). Visualizing large, heterogeneous data in hybrid-reality environments. IEEE Computer Graphics and Applications, 33(4), 38-48.

Redden, J. (2017, December 7). Six ways (and counting) that big data systems are harming society. The Conversation. https://theconversation.com/six-ways-and-counting-that-big-data-systems-are-harming-society-88660

Rosenblum, M. R. (2011). E-Verify: Strengths, weaknesses, and proposals for reform. Migration Policy Institute. https://www.migrationpolicy.org/research/e-verify-strengths-weaknesses-and-proposals-reform

Singh, P., Dwivedi, Y. K., Kahlon, K. S., Sawhney, R. S., Alalwan, A. A., & Rana, N. P. (2020). Smart monitoring and controlling of government policies using social media and cloud computing. Information Systems Frontiers, 22(2), 315-337.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. Journal of Business Research, 70, 263-286.

Stumpf, J. P. (2011). Getting to work: Why nobody cares about E-Verify (and why they should). UC Irvine Law Review, 2, 381.

Taylor, J. (2020). NSW is unable to use Covidsafe app's data for contact tracing. The Guardian. https://www.theguardian.com/australia-news/2020/may/19/nsw-and-victoria-are-unable-to-use-covidsafe-apps-data-for-contact-tracing

Vaithianathan, R., Ryan, M., Anchugina, N., Selvey, L., Dare, T., & Brown, A. (2020). Digital contact tracing for COVID-19: A primer for policymakers. https://csda.aut.ac.nz/__data/assets/pdf_file/0009/382743/PolicyPrimerFINAL.pdf

van den Broek, T. A., & van Veenstra, A. F. (2015). Modes of governance in inter-organizational data collaborations. Paper presented at the European Conference on Information Systems.

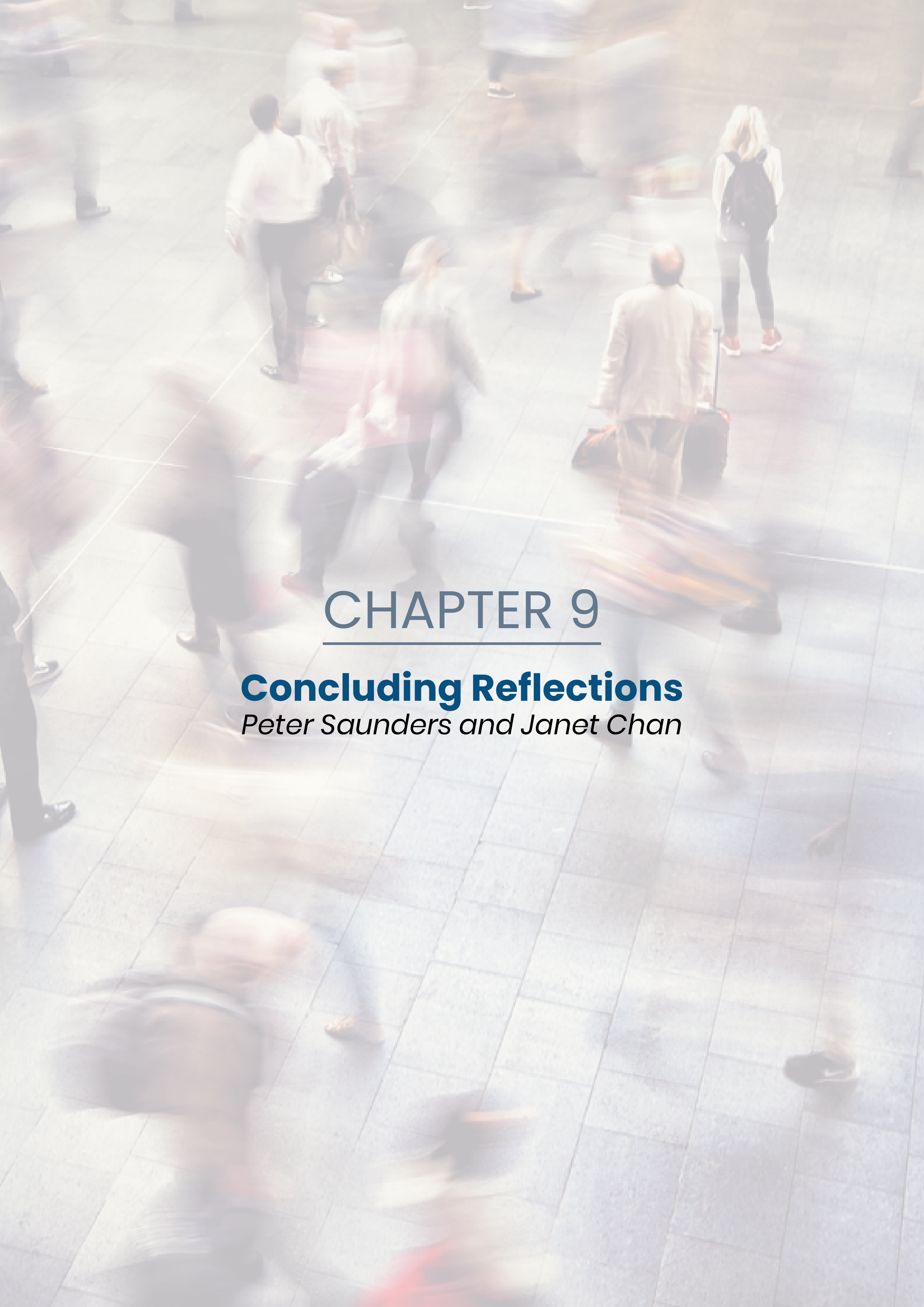
Vatsalan, D., Christen, P., & Verykios, V. S. (2013). A taxonomy of privacy-preserving record linkage techniques. Information Systems, 38(6), 946-969.

Wang, T., Bhuiyan, M. Z. A., Wang, G., Rahman, M. A., Wu, J., & Cao, J. (2018). Big data reduction for a smart city's critical infrastructural health monitoring. IEEE Communications Magazine, 56(3), 128-133.

Williamson, S. (2019). How to fix bias in big data and artificial intelligence. https://thriveglobal.com/stories/how-to-fix-bias-in-big-data-and-artificial-intelligence/

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big data and cloud computing: Innovation opportunities and challenges. International Journal of Digital Earth, 10(1), 13-53.

Zhai, Y., Ong, Y.-S., & Tsang, I. W. (2014). The Emerging "Big Dimensionality." IEEE Computational Intelligence Magazine, 9(3), 14-26.



CHAPTER 9

Concluding Reflections

Peter Saunders and Janet Chan

Concluding Reflections

The preceding chapters provide compelling evidence that the umbrella term “big data” encompasses a variety of major developments that have the potential to fundamentally alter how social policy is formulated, designed, implemented, monitored, evaluated, improved and experienced. It thus offers a unique opportunity to improve lives and wellbeing—often of those most susceptible to the social risks associated with economic progress. It can also contribute to a better understanding of the nature and determinants of social problems, to more efficient uses of economic and social resources, to improved fiscal sustainability and to the legitimacy of state-sanctioned social interventions. Australian social policy is positioned at the cusp of these innovative and exciting opportunities, and most of the benefits of big data lie ahead. Getting there will require an acknowledgement of past failings, the introduction of appropriate adjustments and a willingness to embrace and put into practice the many actions needed to reap the benefits of big data. It will also require acceptance of the risks involved in many big data initiatives, protection of those most vulnerable to adverse impacts, and above all, recognition that the embrace of big data must not involve compromising the underlying objectives of social policy: fairness, inclusion and security.

Big data is not going away. The pace of change is unrelenting, adding to the imperative to develop a framework to shape Australian social policy so that it maintains its distinctive approach without compromising the need for adequate coverage and provision in times of need and equity of treatment of all groups at all times. Experience to date has been mixed—as revealed by the evidence presented in several chapters—but important lessons have been learnt that, if heeded, offer the prospect of a better future. But this will only happen if the responses are open-minded in their conception, appropriate in their specifics and feasible in their implementation.

One theme that emerges from the material presented here is that, despite the many potential benefits of the application of big data to social policy, there has been a tendency for proponents to exaggerate its positive impacts and downplay its potential to produce adverse effects, intended or unintended, often on those most vulnerable. This group has much to gain from the appropriate application of big data to social policy, but only when ethical principles, transparency, trust and open cooperation operate within a governance framework that protects their privacy, rights and independence. The overall calculus must include a broad, inclusive, evidence-based approach that identifies, quantifies and balances the benefits and the risks. It must respond effectively to a range of concerns about data quality, including the errors associated with incomplete coverage (or duplication), self-selection biases and measurement errors, all of which make data linkage particularly prone to compounded errors and inconsistencies. Above all, it must ensure that protecting the vulnerable from the adverse effects associated with errors of design or implementation is accorded paramount importance.

This book draws on a range of big data initiatives in the social sphere to support this endeavour, focusing on the implications for Australia of international practice and the lessons learnt from a number of domestic applications.

The findings contribute to the important task of drawing up a balance sheet of benefits and risks that will guide future applications by drawing on past experience to establish a “best practice” protocol, or at least, to highlight instances to replicate or avoid. The approach of the authors is best described as open-minded but cautious: open-minded in the sense of canvassing the many examples where big data can produce better outcomes, but cautious in its insistence that assessment should be based on evidence informed by conceptual clarity and theoretical reasoning.

As awareness of the inter-connected nature of many social problems has grown, so has the potential for them to become entrenched and/or transmitted intergenerationally. Youth unemployment can lead to physical and mental health problems that have lasting exclusionary effects, not only on those directly affected but on their children; inequalities based on class, gender and ethnicity erode the sense of community that is an important determinant of the social participation that helps generate economic success and social inclusion; alienation and lack of trust in government can undermine people’s acceptance of the regulations needed to promote social cohesion, develop social infrastructure and underpin systems of social support. Transparency is a critical ingredient of the cooperative approach that is needed, but all too often is lacking currently.

Addressing these problems requires generating data that can help identify the key underlying causal and consequential linkages in order to better understand how they determine the intermediate actions that help produce final outcomes. If the underlying problems are multidimensional and inter-connected (as the evidence suggests many are), big data is needed to help isolate key transmission mechanisms in order to identify potentially effective intervention points. It can improve our understanding of the factors driving social distress and provide stakeholders with an important resource to use to combat the flow-on consequences. This can result in better understandings, better policies, better monitoring, better outcomes and better lives.

However, these achievements are not guaranteed. They depend crucially on the nature of the big data itself (including the quality of the data and algorithms used), its legitimacy (or social licence) as an accepted tool of investigation and response, the trust in those overseeing the process, its transparency, the intended aims and the instruments employed to get there, its ability to protect human rights and attract a high degree of public acceptance that the overall process is ethical, fair, transparent and open to challenge. There are positive and negative examples of all of these aspects in the foregoing chapters, which together reinforce the point that benefits are not achieved automatically, even when well-intentioned and eagerly anticipated.

This is partly because of the risks involved, of which there are many. One of these that is highlighted in several contributions is that the benefits are often greatly exaggerated by those supporting big data initiatives. This is apparent from the views expressed by the experts interviewed in Chapter 4, who cite many advantages of big data but acknowledge few downsides, despite the fact that the Robodebt fiasco was unravelling at the time. The Robodebt experience itself cautions against placing too much reliance on big data interventions always going to plan. In this case, the failings resulted from multiple factors, including the lack of human oversight of automated letters to service clients, a lack of respect for the rights of those who received them, the lack of understanding of the distinction between and consequences of false positives and false negatives, an inflexible and under-resourced complaints mechanism and the shifting of onus of proof onto those charged with “offences” that were derived from flawed calculations and often virtually impossible to refute. The contrast between the optimism displayed by the experts interviewed in the first half of Chapter 4 and the dismal realities of the Robodebt experience in the latter half is stark and disturbing.

A second concern that emerges at several places is the lack of transparency of the process used when translating the resource (big data) into improved outcomes (through policy interventions). This lack relates both to the data itself and to the algorithms used to transpose it. When data from different sources collected for different purposes by different agencies is merged or linked, there is a high probability of introducing errors that may be magnified during the analysis phase. Interpretation must be guided by caution, particularly when being used to design new policy reforms. Added to any errors contained in the source data, these dangers can only be avoided through close collaboration between the participating agencies, but this is another area where progress has been inadequate to date. Despite being a core advantage of many big data initiatives, there was a reference in Chapter 4 to an “organisational resistance to data sharing” that was attributed partly to concerns over privacy and consent issues and partly to fears that the data provided might affect future autonomy and funding. This lack of trust in the process produced an unwillingness to share data, even within mainline agencies of different levels of government.

This highlights a more general problem raised in several places: the lack of a comprehensive governance framework to guide big data initiatives, including effective safeguards against misuse and a complaints mechanism with the power to prevent reoccurrence, and, where appropriate, to punish miscreants. Such a framework cannot be imposed from above but should result from a process of open, trust-based consultation that engages with all relevant stakeholder groups (including users). It must be implemented from the outset, not as an afterthought or as a process for seeking approval to key decisions already taken. This remains absent in Australia, as it was in New Zealand, where it was a factor identified as detracting from the otherwise generally well-regarded NZ Social Investment Approach.

Another set of concerns voiced by many of those consulted relates to privacy issues and the related question of respecting the rights of those who provide (and are often surveilled by) big data and datafication processes. Again, the failure to communicate effectively is partly responsible for this being of concern, since it would have emerged and been resolved had there been open chains of communication. There weren't. The implied disrespect for those at the receiving end contributed to a lack of trust that is another factor that emerges in several places as being of concern. Without widespread trust, the legitimacy of big data is brought into question, raising challenges for the social licence that underpins its viability.

A key component of trust is transparency. This is another failing that has multiple dimensions. There is the opacity of the algorithms used in some applications, and a more general lack of transparency over data integrity, data security and data use. Concerns were expressed over the unauthorised use of linked data, and (in both Australian and New Zealand case studies) about the lack of consultation with, and respect paid to the concerns of Indigenous groups and Indigenous issues. Although not an issue of transparency per se, the failure to take adequate account of unintended effects added to the opaqueness surrounding the whole process, while undermining trust in its practical operation.

This discussion highlights the key requirements of best practice when applying big data to social policy. A key consideration here is to ensure that the goals of social policy: social justice, inclusion and economic security for all are not sacrificed in the search for new ways of getting there. There are already many examples where this has occurred, but this is not inevitable if the benefits and risks are balanced. Even the worst cases of big data misuse contain elements that, if applied differently, could have produced better outcomes. Those responsible for the Robodebt fiasco were ultimately brought to account, even if it involved the concerted and sustained efforts of a number of groups. The focus of social policy is on addressing the adverse individual and social consequences of external and induced shocks that restrict people's capabilities and opportunities. These shocks are often not random but are felt most acutely by those already disadvantaged, adding to their vulnerability. This explains why the role of big data in social policy must pay heed to its impact on vulnerable people.

The essential ingredients of this approach include appropriate system-wide characteristics such as adequate governance arrangements and the infrastructure (broadly defined) to support relevant activity. The former is required to establish the trust, transparency, security and privacy that underpin widespread consent and social licence. The latter is needed to ensure that the technology and skills are available to design, implement, use, manipulate, match and interpret relevant data while recognising their limitations and the potential for errors. Accountability is another key requirement, with major stakeholders accountable to each other and to the general public, but also to specific groups, including Aboriginal and Torres Strait Islander communities and other vulnerable groups in the population. Finally, all participants need to have realistic expectations about what big data can achieve and an awareness of the likely obstacles to success that will make its application grounded in knowledge and experience.

Important initiatives are underway that will promote these goals. They include the Integrated Data Infrastructure (IDI) initiative in New Zealand as part of its social investment approach, and the Multi-Agency Data Integration Project (MADIP) being led in Australia by the Australian Bureau of Statistics. The "Five Safes" model discussed in Chapters 5 and 8 should be a central feature of these developments, in order to achieve safe projects, safe data, safe outputs, safe people and safe settings. This would provide an easily communicable ethically based "best practice" framework that would assist big data initiatives to achieve their objectives, but also help to build community trust in the overall approach and promote other important goals such as privacy, accountability and the protection of human rights.

The chapters that discuss the nature and impact of data linkage initiatives in three important areas of Australian social policy—income support for the unemployed, improving the wellbeing of vulnerable children, and place-based initiatives for families and children—identify three core components of effective data linkage programs in these areas. The first is the development of appropriate technical expertise, the second, a detailed understanding of social policy and practice, and the third, building trust through engagement with the community, data custodians and other stakeholders. These are straightforward evidence-based practical requirements that provide a valuable starting point for ensuring a better future for big data in the social policy arena.

The conclusions drawn in Chapter 7 on what the earlier chapters imply about the ethical nature of big data highlight this as a major shortcoming. The review identified a significant gap between widely accepted ethical principles that apply in relation to legal and social policy and actual practice. The authors note that issues such as human rights, the rule of law, freedom, dignity and accountability are rarely considered in the design and implementation of big data programs or initiatives. This failing is problematic in and of itself, but can also arise from the inadequate treatment of other issues, including technical deficiencies in data quality, data security, linkage accuracy and inadequate skills and capacities, all of which may give rise to ethical problems such as fairness, due process and social justice. These issues are core principles of social policy, and the need for their application and enforcement to big data in this area is of utmost importance.

Chapter 8 addresses issues surrounding the importance of infrastructure. This also emerged as a concern in earlier chapters, although there was apparent confusion in distinguishing between technical and infrastructure issues among some interviewees. Two important points were made. The first was that appropriate digital infrastructure is critical to the success of data-driven solutions for decision making. The second is that different technical considerations are required at different stages of the data analysis pipeline. Choices made about what infrastructure to develop can also affect the outcomes achieved, which implies that infrastructure (like ethics) must be a core component of big data developments, not an “add-on,” as they have both tended to be treated in the past.

ACADEMY OF
THE SOCIAL SCIENCES
IN AUSTRALIA

Academy of the Social Sciences in Australia Inc.

GPO Box 1956, Canberra, ACT, 2601
info@socialsciences.org.au
02 6249 1788
www.socialsciences.org.au