

# Secondary use of government-held personal data: ethics, law and governance in Aotearoa, New Zealand and Australia

A workshop at the University of Auckland and online, 28 June 2022

Assoc. Prof. Barry Milne and Dr. Megan Pictor co-hosted this event, supported by a Dyason Fellowship (Pictor) and Te Rourou Tātaritanga, The Informatics for Social Services and Wellbeing Programme funded by MBIE Endeavour Grant, ref 62506 ENDRP (Milne).

This workshop aimed to share trans-Tasman knowledge about secondary use of government-held personal data around three priority issues:

- trust and transparency;
- group data harms; and
- Indigenous data sovereignty and data ownership.<sup>1</sup>

There were six speakers, of whom two— one from each of Australia and New Zealand— spoke to each priority issue. After each pair of presentations, a whole-group discussion was conducted under Chatham House rules.

**This report reflects the knowledge shared over the event. It may be used to inform further work on secondary use of data. The convenors thank the participants for their contributions and willingness to engage in discussion.**

## Indigenous data sovereignty and data ownership<sup>2</sup>

Assoc. Profs. Māui Hudson (University of Waikato) and Luke Burchill (Mayo Clinic/University of Melbourne)

### Presentations

A/Prof. **Maui Hudson** told attendees that the issue of secondary use of government data is that it does not 'sit within tribal hands.' Indigenous data sovereignty is about rights and interests, but holistically, data control is intimated in the notion of Indigenous data sovereignty. Therefore, secondary use of government data doesn't easily fit within this control paradigm and will rely on other mechanisms to affect Indigenous rights and interests. Other mechanisms might include the CARE (collective benefit, authority to control, responsibility, and ethics) and FAIR principles (findable, accessible, interoperable, reusable). Another mechanism of recognising Indigenous rights and interests is exemplified in the Yoorrook Justice Commission<sup>3</sup> (Victoria, Australia), which has adopted the Maiam nayri Wingara Australian Data Sovereignty Principles. However, principles are not sufficient. There is a need for various tools and mechanisms to enable enactment of principles over the data life-cycle. Ngā Tikanga Paihere<sup>4</sup> is one tool designed to share traditional knowledge to manage Māori data over its lifecycle. Hudson observed that this tool could 'map quite well' to the Five Safes<sup>5</sup> in a 'culturally

<sup>1</sup> Nine stakeholders were asked to identify and then prioritise key issues with respect to the topic. The top three themes were selected for the workshop.

<sup>2</sup> The Indigenous data discussion is recounted first in this report, responding to experts' opinions that Indigenous data discussion be framed in upstream, mainstream governance. We seek to place this discussion prevalently and centrally.

<sup>3</sup> <https://yoorrookjusticecommission.org.au>

<sup>4</sup> <https://data.govt.nz/assets/data-ethics/Nga-Tikanga/Nga-Tikanga-Paihere-Guidelines-December-2020.pdf>

<sup>5</sup> See, e.g. <https://blog.ons.gov.uk/tag/5-safes-framework/>; <https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/>; <https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework>

grounded manner.’ Notably, the cultural grounding occurred independent of the Five Safes and was then mapped onto them, rather than the other way around.

In addition to Indigenous data governance, there is reason to consider Indigenous governance and data. Hudson suggested that Indigenous traditions and thinking tend to consider the collective and, therefore, might be a better fit than Western traditions for data governance of common resources or for collective benefit. This prompted a broader consideration, ‘how to challenge existing systems’. Challenging systems creates an opportunity to improve governance. This is important to counter the increased risk of perpetuating issues and inequality through data systems and AI. Academics, regulators and designers need to consider how to build protections and challenges to data governance systems. Continued work needs to address standards<sup>6</sup> and continue to define Indigenous rights and interests in data. Further, this conversation requires a focus on benefits from data not only risks in data.

A/Prof. **Luke Burchill** began by sharing his story of growing up on Country and tradition of the Yorta Yorta people. From here, he positioned his crucial message, explaining that Indigenous governance issues should be situated in the mainstream rather than in specific areas. His approach to research is to think about the community defined as those with the most to lose and the most to gain. He described the primary concern with research governance is that it is framed in line with Western conventions that govern from the top-down, such as through funding decisions and timelines. This reduces the ability of researchers to respond to the needs or wishes of community, except incidentally. He suggests that there is a missing link in research to place, people and relationships, but that Indigenous governance can address that, and this link should be important to governance generally.

Burchill noted that in Australia, most Indigenous research is done by non-Indigenous persons at non-Indigenous institutions. There are fewer Indigenous inputs into the research process down the research pipeline, especially and essentially, at the analysis and meaning-making stage. This stark reflection was a strong paradox to Burchill’s starting point, where he acknowledged the Aboriginal practice and expertise in meaning-making and story-telling. Quoting Maggie Walter he highlighted, ‘[t]he primary problematic is that the Indigenous ways of seeing the worlds are not doing the shaping.’ In the final part of his presentation he introduced his conception of ‘culturally adaptive governance’ which blends Indigenous governance, adaptive governance and critical allyship. The approach would facilitate the reshaping already described, albeit in a health research governance context.

### Discussion highlights

- History is an essential starting point in connection with people and place in data governance.
- Indigenous self-determination is essential, as are notions of equity, diversity and inclusion. Each has ability to shape and enhance academic conversation in this area but should be supported by principles and process mechanisms.
- Contextualising mainstream data governance principles in Indigenous context may be insufficient to meet the needs of, as well as bring benefits to, Indigenous people. An important part of this discussion was that the pace of change of technology and the pace of Indigenous rights development are diverging. Although incremental change is seeking to improve Indigenous rights, greater effort is required to ensure that Indigenous right not only improve but are brought in alignment with governance efforts.

<sup>6</sup> See e.g., <https://standards.ieee.org/ieee/2890/10318/>

- *There are serious issues for increasing Indigenous capacity to lead the work. Indigenous and First Nation peoples tend not to be supported nor rewarded for leading Indigenous work for non-Indigenous people and groups. It can also be harmful and dangerous to individuals.*
- *There are different issues with respect to data governance in NZ and Australia but both have enduring challenges. This emphasises the connection to people and place. Where there are social failings including racism and inequity, these should be addressed at a policy level, acknowledging the risk that AI will exacerbate and cannot mend data governance systems that are already hindered.*
- *An important point was made about building systems and selecting data that recognises strengths and areas that are performing not building systems on deficits. e.g. there is a problem is 'disease is rewarded'.*

### Future work

- *Placing Indigenous governance in mainstream health governance.*
- *Intersectionality, increasing representation of Indigenous people in government and data governance; radical allyship.*
- *Increase awareness of the dangers for Indigenous peoples, including as researchers in mainstream health and health data research.*
- *Investigate standards for strength-based governance and use of data for Indigenous peoples and collectives.*
- *Investigate standards and mechanisms of challenge in data systems.*
- *Relevant to all standards: an investigation into contextually grounding data governance in Indigenous perspectives and approaches. Part of this should consider what is lost or missed if we begin with a Five Safes approach compared to where an unincumbered Indigenous rights-approach might lead?*
- *Consider what can Indigenous approaches to governance bring to data governance generally?*

### Transparency and trust

Assoc. Prof. Angela Ballantyne (University of Otago) and Dr. Megan Pricor (University of Melbourne)

### Presentations

A/Prof. **Angela Ballantyne** spoke about the need for transparency and trust in data governance. She highlighted the clarifying potential which arises from recognising data controls as either a personal authorisation mechanism or as an instrument of public reason/interest controls. Explaining that social licence relied on transparency and trust, she described the concern that personal authorisation mechanisms, especially consent, are a 'thin' form of authorisation for secondary use. This could trigger a preference for a more substantive, collective value approach to secondary use mandate through public interest. Ballantyne raised serious concerns about the publicly conceived levels of trust and competence (referring to the Edelman Trust Barometer 2022)<sup>7</sup> in government in NZ. Low levels of trust impact the social licence concerning secondary use of government data. She highlighted the importance of improving the public perception of competence (in addition to and in slight preference to 'trust'<sup>8</sup>) and building an ongoing approach to obtaining and maintaining social licence. This was in part predicated upon the importance of public transparency. The data, commonly both about the public and funded by the public, should be available to the public. Importantly this data should be accessible and not 'buried', unable to be found by those who seek to find it. She highlighted

<sup>7</sup> <https://www.edelman.com/trust/2022-trust-barometer>

<sup>8</sup> This is in line with the literature on trustworthy governance, urging governments and decision makers to focus on trustworthiness rather than obtaining trust, including the prominent work of Onora O'Niell.

the NZ Futures data tool (although the organisation no longer exists)<sup>9</sup> as a valuable instrument to guide transparency.

Dr. **Megan Pricor**'s presentation focused on My Health Record in Australia, highlighting that whilst the tool is underutilised, it holds great promise for empowerment by giving individuals straightforward access to their health information for the first time. My Health Record was developed to address health information fragmentation and after an initially low uptake, the shift to an opt-out enrolment means that approximately 90% of Australians have a record. The individual can see all the data held about them and has a range of options to manage their profile. In general, the system has challenged people's trust in government, including with respect to concerns that people's health data could be accessed by enforcement agencies without a court order, or that health record data might be sold. The design of My Health Record, whereby all of a person's treating health professionals can see their whole record by default, has also caused unease.

The secondary use framework for My Health Record responds to most of the issues identified in the literature.<sup>10</sup> Pricor suggested including elements of a Dynamic Consent approach within the My Health Record platform, including options to select one's own preferences about data use in broad areas of research and to provide consent for the use of identifiable data for research. The My Health Record platform could thus provide for a 'communication channel' about data flowing for secondary use – noting that two-way communication between researchers and data subjects is a key element of the Dynamic Consent approach. In line with Ballantyne's presentation noting social licence, Pricor ended considering the Wellcome Trust model for Learning Data Governance, suggesting that in addition to heavy front-end governance models, promulgating a learning approach via incorporating citizen panels to increase data decision-making over time would be of likely benefit.

### Discussion highlights

- *If data subjects are less engaged with data governance instruments (e.g. My Health Record preferences) there is good reason to consider the collective protection of those groups' data.*
- *Where does trust lie? The group expressed a need to understand whether a government was subject to higher public expectations of ethical positioning and competence compared with companies. If this makes it easier for companies this may not necessarily mean that companies are preferred developers but raises important avenues of research with respect to perception, expectations, preferences and licences. Additionally, how does this relate to the relationship between income and trust which seem to have a correlated relationship?*
- *Where do rights lie? There are reasons to identify data as either personal or collective and both. The framing of personal data in the consumer context, although sometimes problematic, supports the recognition that those who benefit from data are put on notice of their responsibilities to those who are more vulnerable to its use. There was a particular concern that if collective rights are preferred that minority concerns might be lost.*
- *The issue of transparency arose with respect to purposes of government data use; citizens should be aware of what information governments hold so that they can have confidence in its use.*
- *The issue of correction: to what extent should patients be allowed to change information about themselves?*

<sup>9</sup> As this is no longer available through NZ Data futures, see the decision wheel reproduced, e.g., <https://openqovasia.com/data-futures-partnership-in-new-zealand-issues-guidelines-for-organisations-to-develop-social-license-for-data-use/>

<sup>10</sup> including no access by insurance agencies, no use to determine eligibility for benefits; the data will not leave Australia; aligns with current guidelines including Five Safes and there will be a public register for requests to access the system.

- *There was some discussion about how community panels are used in governance.*
- *An overall concern emerged about how a government or society might or should improve trust/trustworthiness? How does this relate from a shift from ‘trust us’ to ‘you can trust us’ (trustworthiness)?*

### Future work

- *Can both individual and collective rights lie in data?*
- *Should/ when should individual authorisations be preferred over collective approaches? When is the consent model valuable? When should it be preferred over collective models? When is public interest most important?*
- *Should/how should community panels be used in decision making?*
- *How can trustworthiness be improved by a government?*
- *Should individuals be able to change information held about them?*
- *How can information transparency be supported so as not to bury the most important information from publics and actively enable accessibility?*

### Group data harms

*Mr. Andrew Sporle (University of Auckland) and Prof. Jeannie Paterson (University of Melbourne)*

### Presentations

*Mr Andrew Sporle discussed the potential issues of group data harms from an NZ perspective. He asked ‘How can we develop a better understanding of group harms’ for the purpose of doing better? Sporle’s discussion begun by recognising the extent and size of government data. He explained that the information is an excellent statistical resource, very complex, large and diverse, but this means it is decreasingly observable and accessible. The impact of secondary data, collected over time from different sources, is that users are ‘distant’ from data subjects and one-way data flows begins to feel like surveillance – even when it’s not. Additionally, government data is ‘siloe’d’; the knowledge of its existence is the first barrier to use.*

*Once knowledge of data sets is obtained there should be consideration of the purpose for which the initial data was captured and whether it is appropriate data for secondary use. He explains this through the example of a ‘potential demographic dividend’ – the NZ population of European descent is skewed towards older people so that in the near future the total number of people over the age of 65 will be larger than the total number of people under 50 but this is not true of the Māori population. Ignorance of this will likely result in inappropriate use of the data – and missed opportunities.*

*Data use requires more accurate understanding of the data held and its quality –and further, of collectives it seemingly imparts knowledge about. When discussing data harms we must identify the collectives first, and consider the accuracy of the claims the data seeks to import. Harms can arise or exist at all stages of the data life cycle. Further, analysis of the quality and purpose should be reported to contextualise results. Contextualising the quality of data opens pathways to change the data collection and use. Notably minorities are likely overrepresented in data this influences the stories that data can tell. “The state knows more about disempowered people than they do about empowered people”. The current approach is to focus on deficit data rather than success stories. He asks, how can you use deficit data to lead pathways to success? Equally importantly, how do the pathways represent what is wanted by Māori people? The metrics should be valuable to the communities they seek to represent.*

*Prof. Jeannie Paterson focussed on secondary use of government data with respect to AI, such as using statistical models to predict future behaviour. Without discounting the benefits, one risk in such processes is algorithmic bias affecting in particular groups already experiencing intersectional disadvantage, and discrimination. Paterson argued that the risk is real and the solutions lie primarily not in mathematics but in understanding the role of ethics, values and policy in deciding what use to make of AI, data analytics or machine learning. This also requires a diversity of skills, perspectives, and experiences in decisionmakers.*

*Algorithmic bias arises in a variety of ways. These include data collection and handling practices, which may result in the underrepresentation or misrepresentation of certain social groups. The common feature is that it arises from human decisions, prejudices or behaviours. Humans already make biased decisions. The concern with algorithmic bias is that it arises from processes operating at scale, at great speed and often opaquely, which means amplifying and replicating bias.*

*Identifying and responding to the risk of harms to groups through algorithmic bias is not straightforward. Algorithmic bias may be difficult to identify, particularly where using proprietary or complex machine learning algorithms. It may be unclear on what basis a prediction is made, and whether it represents a new insight on a population or group, or a biased outcome. Sometimes the issue may be resolved by ensuring greater diversity in the training data. But typically, this does not solve the problem of situations where otherwise neglected populations also present differently to the majority, and yet with small numbers they may exit as outliers in the data. It is sometimes suggested that concerns about algorithmic bias can be met by removing gender race etc as data points or giving them different weightings. But the solution is undermined by the presence of proxies for protected attributes in the data sets. There may be a necessary trade-off between predictive accuracy for the majority and fairness for different groups within the data set.*

*It is sometimes said a human in the loop is the answer to concerns about bias in algorithmic predictive or decision-making tools. But the approach deflects attention from the systematic risk of harm in such approaches, and also the tendency of humans to accept what is presented via automated processes as objective and neutral. Part of the response lies in rigorous governance processes around the use of data for automated decision-making. Responsible AI, encapsulates a suite of safeguards and processes to address bias (although they cannot solve it), including auditing, testing and reviewing outcomes. Transparency is key throughout all processes. Another aspect of responsible AI is the importance of inclusivity in design and ongoing use, meaning involving people affected in the initial design and in development of AI.*

*Paterson argues for a human-centred approach to address algorithmic bias as they are going to be resolved by policy decisions. Bias arises from humans, and this requires social policy decisions. e.g. 1, if we do not have relevant data, we need policy decisions about what criteria are relevant to the decision. e.g. 2, if we have over-represented data, the process to address this is asking why this group is overrepresented and then consider the response in light of the social factors.*

### Discussion highlights

- *There was a discussion of transparency and oversight in public sector decision-making. The first step is to map uses and improve clarity around use of AI in government. There was reference to the EUAI Act which combines governance oversight with transparency. In general, there is a tendency for transparency to divert attention. The counterpart to transparency of data use is transparency around the oversight.*
- *Much discussion concerned a human-centred approach to AI. How can we begin to see the links between human decisions and data sets? Firstly, users of technology might tend to overestimate the ability of*

technology to fix AI. Secondly, transparency is a precondition to understanding uses and data of AI.

Thirdly, transparency requires oversight of the technology, and here we begin to open the way for links to human decision-making. Fourthly, it is necessary to assess the mechanisms of the governance, by documenting governance and considering the contestability of how decisions are made (in all aspects).

- AI requires contestability of data (Robodebt<sup>11</sup> and bad algorithmic decision-making). There are important legal mechanisms for contestability as a prompt for good governance (NSW Ombudsman has made a good example of contestability).
- Ethical use of AI decision-making and humans. The relationship between humans and AI is complicated. Sometimes humans are 'just as biased' and may be unaware. 'Human in the loop' is an approach to managing AI. This may result in the human deferring to the AI where they don't understand the basis on which the AI makes the decision or don't have time to engage with the decision at the relevant level. We know that often human panels might be more biased than AI and this is an opportunity for inclusive design. Process decisions are probably able to be resolved by AI, but for access to resources, and interaction with government agencies, humans should be involved. The way that the human is involved in the system or nudged is important! Humans should be involved and reflect on the processes as we begin to use AI.
- Any time AI is used with respect to a collective, that collective should be involved.
- Responsibility and accountability should be considered in the design and transparency decisions. There is a tendency to blame the individual in a scenario with a bad outcome, but accountability means sharing accountability throughout the systems and processes.

### Future work

- Is the reused data fit for the new purpose? This is rarely assessed.
- Data triggers for policy decisions.
- A human-centred approach: how should humans and AI come together?
- What are appropriate data contestability mechanisms? How can we build these into governance systems and law?
- How to integrate processes to facilitate system accountability and share responsibilities across decision-making systems, not only among the human decision-makers?
- Legislative approaches for transparency in AI.

The convenors thank Ms Tess Whitton (University of Melbourne) for her assistance in the preparation of this report.

<sup>11</sup> For instance: <https://www.theguardian.com/australia-news/2022/may/14/a-failure-as-shameful-as-robodebt-leaves-questions-only-a-royal-commission-can-examine>