

## **Data Ethics Governance Workshop - University of Otago, Wellington**

### **Summary of Discussion**

Prepared by: Angela Ballantyne, Maria Stubbe, June Atkinson, Rochelle Style, Ryan Orange

---

**Workshop aim:** To identify the key challenges regarding data ethics and governance from a cross-agency and cross-sector perspective, with a primary focus on large public sector data sets. The relationship between citizens and government is different from that between a consumer and private company, and this 'social contract' framed discussion of the ethical issues.

#### **Challenges**

- Māori Data Sovereignty, Treaty responsibilities and the concerns of Māori in relation to historical and potential future harms arising from data use are currently not well understood or integrated into discussions and planning across the sector.
- Current data use is collapsing many established conceptual, ethical and regulatory boundaries, and this makes data governance challenging (for example 'public versus private sector data', 'research versus operational use').
- The current data ecosystem is fragmented and unstructured, resulting in duplication of governance and review structures, governance gaps and/or conflicts.
- The cycle of government regulation is typically much slower than the average technology cycle, so regulation will always be a blunt tool to try to control developments in data technology.
- Data longevity and future uses of data are unpredictable, making 'informed consent' and robust governance increasingly problematic.
- Controlling data use is becoming increasingly difficult and we should be strategic about where we try to intervene in data pathways and to what extent we attempt to exert control.

#### **The nature of research**

- Rapid changes in the nature of research, the way we use and interrogate information.
- We are moving from a model where personal data is predominantly 'held' by the individual and given to service providers or researchers, to an 'observed environment', where data subjects are under constant observation and the service provider in many cases already has the data. It is then not a question of giving researchers access and permission to use data, but rather attempting to intervene to stop them using data inappropriately.
- There is a data surfeit, and entry costs to some kinds of data research are now low. This raises concerns regarding the expertise of those working with data, the quality of data research

methodologies, and the potential for unwarranted conclusions to stigmatize minority and 'vulnerable' groups and to pervert service delivery.

- A possible end to hypothesis-led research (and even open-ended question-led research) was noted, along with a shift towards using data for service delivery 'because the computer algorithm says so'. There was general concern about unrestrained enthusiasm for numerical data ('just go look in the IDI'); and lack of careful consideration of appropriate research methodologies, including qualitative approaches ('matching methods to research questions').
- Further to this, it was noted that the types of data/datasets held in repositories such as the IDI are currently far from comprehensive, and may themselves be vulnerable to 'selection bias' or 'tainting' in various ways, making critical and robust peer review of research methodologies and questions even more essential.

### **Shifting focus: individual control to collective interests**

- Established ethical frameworks (particularly in relation to health information) focus primarily on individual control of and consent to data use. This framework is becoming difficult to apply to current data practices. Statistics NZ's "5 Safes" ethical framework instead focuses primarily on data security and public benefit (though how the latter should be defined is not a settled issue).
- It was noted that much public discussion focuses on the *risks* of data use and that the *benefits* of data use may be under-emphasised. (By contrast, there was concern that data researchers themselves may not sufficiently consider the ethical risks). Conversely, what are the costs of not using data and relying on other models of research or service delivery?
- Given the volume of data collected about individuals and the capacity to link datasets, the ability to genuinely anonymize data is decreasing. Even when data is de-identified and aggregated, meaning there is less risk to individual privacy, data use and linking of data sets can still result in the identification of collectives and can result in harm or benefit to these groups. It was noted that the current privacy regulation (HIPC) is enabling in that it allows wide discretion for HDECs to approve the use of potentially identifiable health information without patient consent when this is in the public interest.

### **Consent**

- There were conflicting views on the role of consent, with some participants committed to consent as the primary ethical value (e.g. 'the first question has to be, can you get patient consent?'), whereas other participants considered individual consent practically impossible and/or ethically unnecessary in relation to data use for certain purposes.
- It is impossible to know at the point of data collection all future uses the data may be put to.

- There was general scepticism about the value of ‘consent’ that was required as a condition of accessing services (e.g. Microsoft, Facebook, registering with a PHO, the MSD “NGO data for funding” model); and of opt-out consent models (e.g. some public data registries overseas).
- In the absence of individual consent, other mechanisms can be used to involve communities in governance of data use.
- There was consensus that clearly articulated principles need to underpin whatever governance and review structures are developed. Discussion focused on the following ethical issues: a people-centred approach; principles of Māori Data Sovereignty; and trust and transparency.

#### **A people-centred approach: ‘Nothing about us, without us’**

- Data users should ask: Where is the human in the data? What would this data use look like from the data subjects’ perspective?
- Data research, use, and interpretation should avoid a ‘deficit model’ that focuses primarily on the ‘poor performance’ of individuals or groups against some arbitrarily defined metric.
- Data is used to tell stories to help us understand the world. This process is never ethically neutral; narratives always have embedded cultural values and ethical assumptions. We should be especially careful when powerful agents (such as governments) collect and use data about minority groups or vulnerable individuals/sectors of society.
- It was noted that the most vulnerable members of Aotearoa New Zealand often rely on multiple government services, experience the greatest surveillance of their activities, and therefore have the most data produced about them, but also have the least capacity to influence the narratives that are drawn with this data and the resulting policies and practices.
- It is important to have the right people involved in data research interpretation (framing, anchoring, and projection), especially when data analysis is used to inform or drive the provision of government services. Iterative cycles of consultation with relevant stakeholders (including communities) are essential throughout the data collection and research processes.

#### **Kaitiakitanga: Māori Data Sovereignty**

- Kaitiakitanga means guardianship and protection. Māori data sovereignty refers to the idea that Māori data is subject to the rights articulated in the Treaty of Waitangi. Māori data (including data about Māori, and data used to describe Māori collectives, and data about Te Ao Māori) is a living tāonga and should be subject to Māori governance.
- The rights and interests of Māori in relation to data collection and use emerged as important focal themes from the presentations about MDS and the post-colonial context, and it was acknowledged that there was a clear need for people and organizations across the sector

(including many of the workshop participants) to gain a deeper understanding of Maori perspectives, and to take these into account more explicitly.

- What data arrangements (for collection, storage, use and interpretation of data) will best serve the interests of Māori? Who has authority to answer this question?
- If data analysis is to be used for the benefit of 'NZ Society' – who decides how 'public benefit' is defined and how potential benefits and risk should be distributed throughout the community?
- How do we make sure that the interpretation of data *about* Māori is accurate and appropriate (especially when this is being done by non-Māori analysts)?

### **Trust and Transparency**

- Governance structures should improve and maintain transparency and enable a high level of public trust in data governance and use.

#### *Trust: 'Social licence' and 'cultural licence'*

- There was an acknowledgement that government policy regarding data use was moving ahead of the public conversation and 'social licence' (aka 'social contract').
- The 'social licence' refers to the degree to which the public thinks that data use is legitimate. Te Mana Raraunga (the Māori Data Sovereignty Network) has recently proposed the idea of a 'cultural licence' for data use, referring to whether data use is legitimate from the perspective of iwi and Māori Treaty partners.
- Some scepticism was expressed about this concept: How does the 'social licence' legitimise data use? What the public finds 'acceptable' does not always correspond to 'ethical' practice. Can people/groups make informed choices when they have little knowledge about the extent to which data about them is collected, linked, analysed and used for various purposes?

#### *Transparency*

- Transparency emerged as an important theme, particularly in the absence of consent. Transparency is important to support accountability for government data use; facilitate public debate and critique; and to share knowledge (regarding ethical review issues, algorithmic design, research results).
- It is valuable to have the minutes of government data ethics committees publically released (as per HDEC meeting minutes).
- Transparency is important for building a high-trust environment that can support data use.

#### *Data ethics review*

- Review of data use varies across government agencies, and across sectors: Research involving potentially identifiable health information typically goes through independent review by an ethics committee. IDI data use goes through the "5 Safes" checklist and must be approved by

the Government Statistician. Private companies and some government agencies have internal non-transparent data ethics committees (e.g. Google's DeepMind; MSD is working to develop a Privacy, Human Rights and Ethics framework to help guide operational data use and is in the process of establishing an independent Interim Research Ethics Panel to provide advice on research and evaluation while wider conversations about the possibility of a strengthened cross-agency or inter-sector ethics infrastructure play out.)

- There was considerable interest in a developing an inter-sector 'public data use' review agency that could provide *advice* on data use for research, service provision (e.g. predictive risk modelling), linking of government data, and sharing of administrative data with the private sector or entities outside of Aotearoa New Zealand. This review committee should include expertise in data analytics, statistics, observational research methodology, Māori interests (e.g. Te Mana Raraunga), ethics, and community/lay representation. It was acknowledged that it would be impractical to review all data use.
- Accountability: can sanctioning or punishing data misuse act as a future deterrent and contribute to a culture of careful, ethical data use? The idea of a Data Ombudsman, with powers of investigation and redress, was proposed.

#### **Expertise and training:**

A major theme of the workshop was demand for greater expertise across multiple levels.

- Data researchers can learn more about the sources of the data they access (i.e. the context(s) in which it was collected), the power of data to stigmatize groups and the ethical issues involved in data use. We can improve public understanding about how people's personal data is used in both the private and public sectors.
- Māori Data Sovereignty principles should be more systematically and consistently integrated into national debates about data ethics.
- We can focus on training more data researchers from diverse backgrounds, including Māori. This is a key mechanism for ensuring 'nothing about us, without us' and democratising the capacity to use data.
- We need to ensure that all those working with 'big data' understand the methodological limitations inherent in doing research and analysis with this type of data.
- We need to develop tools (including ethical frameworks) for training researchers and agencies about ethical and cultural concerns to consider in relation to data use.

---

## **Actions**

*Please get in touch if you would be interested in contributing to any of these activities.*

- Develop a 2-page framework of key ethical issues in relation to data collection, storage and use for government agencies. (Angela, Tim, Hamish)
- In response to the workshop, we are considering running a University of Otago, Wellington Public Health Summer School day long work shop on data ethics; keep an eye on this page <http://www.otago.ac.nz/wellington/departments/publichealth/summerschool/> for details (June, Angela, Rochelle, Maria).
- Work on transparency as an ethical tool in relation to data (Angela and Rochelle).
- Continue to discuss the role of a government wide Data Ethics Review Committee.
- Participants were encouraged to contribute to the upcoming consultation on the Statistics Act 1975 Review.

## **Links to resources mentioned in discussion**

IEEE Advancing technology for humanity “Code of Ethics”

<http://www.ieee.org/about/corporate/governance/p7-8.html>

Digital Economy Act 2017 (UK) <http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted>

Te Mana Raraunga <http://www.maramatanga.co.nz/news-events/news/te-mana-raraunga-m-ori-data-sovereignty-network>

Statistics NZ - “How do we keep IDI data safe?”

[http://www.stats.govt.nz/browse\\_for\\_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx](http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx)

## **Definitions**

Data subject is the individual to whom the data refers.

Data producer is the entity who has collected, compiled, curated and/or stored the data.

Data recipient is the entity who receives access to the data for a secondary use.

Stakeholders are those people or entities that have an interest in the data.

Communities (or communities of interest) are the groups comprising the data subjects; although not every member of the community may be a data subject. Communities are one type of stakeholder. They may be defined in various ways (e.g. by geography, education, socio-economic status, ethnicity, religion, culture, service use, health status or disability), and one dataset might relate to multiple communities. Communities may experience benefits, burdens and risks relating to the sharing and use of data, and these can be distinct from the benefits, burdens and risks experienced by individual data subjects.

**Organisations represented at the Workshop**

1. Capital & Coast DHB (Breast Cancer Data, Research Office)
2. Data Futures Partnership
3. Lakes District Health Board
4. Ministry of Business, Innovation and Employment
5. Ministry of Health (HDECs, NZ Health Survey, Data Analytics)
6. Ministry of Social Development
7. Office of the Privacy Commissioner
8. Oranga Tamariki, Ministry for Vulnerable Children
9. Statistics NZ
10. Te Mana Rauranga
11. University of Auckland (NETwork! Research Program, Philosophy)
12. University of Otago (Research Ethics Committees, Virtual Health Information Network,  
Departments of: Public Health, Philosophy, Primary Health Care & General Practice)
13. Waikato University (Ethics Committee)