Mana motuhake ā-raraunga: datafication and social science research in Aotearoa

Tahu Kukutai & Donna Cormack

Routledge
Taylor & Francis Group

EDITORIAL

Mana motuhake ā-raraunga: datafication and social science research in Aotearoa

Social practices, interactions and relations are increasingly being turned into data, driven by technologies that enable new methods of data accumulation, digitisation, integration and manipulation (Taylor and Broeders 2015; Couldry and Yu 2018; Mann and Daly 2019). This 'datafication' is not only transforming social relations, it is also reshaping social science through the generation of datasets that open up new avenues for research and methodological development (van Dijck 2014) and generating debate about the relationships between data, digitalisation, ethics and human wellbeing (boyd and Crawford 2012; Mittelstadt and Floridi 2016; O'Neil 2016; Eubanks 2018).

Datafication is playing out in Aotearoa New Zealand with rapid developments in datasharing, linkage and the integration of large datasets for operational and research purposes. While the Government has distanced itself from its predecessor's social investment approach (Boston and Gill 2017), there remains a focus on data-driven decision-making that includes the use of integrated data, operational algorithms and predictive risk modelling (Stats NZ 2018). The flagship of integrated data is the Integrated Data Infrastructure (IDI), a major research database managed by Stats NZ that links microdata about individuals and households from the census, surveys, government agencies and an increasing number of NGOs. The deidentified data are available under the 'five safes' framework for research that is 'for the public good' and that 'improves) the outcomes of New Zealanders' (Stats NZ 2019a). Recent research includes projects that focus on the causes and consequences of criminal activities; the health consequences and costs of child poverty; and the development of models to predict suicide and self-harm risk (Stats NZ 2019a). Applications to use IDI data are reviewed internally by Stats NZ. Their process does not require approval from an ethical review committee, nor external peer review from subject experts (Gulliver et al. 2018).[1] Outside of Stats NZ, there is no standard approach to the ethical use of IDI data across the education and research sectors. Health and Disability Ethics Committees, established under the New Zealand Public Health and Disability Act to ensure that health and disability research meets ethical standards, do not currently require ethical approval for the use of IDI data in health research projects. Some universities require that researchers submit ethics approval to use secondary datasets, such as the IDI, while others do not. The IDI is also used by researchers working in government departments, think tanks and independent consulting companies that do not have internal ethics processes.

Social licence and Census 2018

For some researchers, the absence of ethical oversight over the IDI might be seen to absolve them from engaging with issues arising from datafication that include consent, trust, benefits, harm and risk. Our view is that the lack of mandatory ethics should prompt more, not less, scrutiny and reflection about how integrated data are produced and used. Take, for example, the 'social licence' that Stats NZ depends on in order to undertake its data integration activities (Nielsen 2018). Gulliver et al. (2018) have traced the concept of social licence to the work of sociologist Everett Hughes who 'explored the conditions under which society was prepared to afford professions permission to adopt practices that violate accepted social norms without incurring social sanction' (p. 59). Social licence has been most widely used in extractive sectors, such as mining and forestry, where industry practices can result in harm to the environment or communities. In those contexts, organisations can be said to have a social licence to operate when their conduct and practices meet societal or community expectations. Stats NZ defines its social licence as the permission it has to make decisions about the management and use of the public's data *without sanction* (emphasis added, Nielsen 2018).

One might ask, to what extent does Stats NZ have a social licence to integrate individual and household data and make it available for statistical and research purposes, including social science research? Recent survey research commissioned by the agency suggests that its social licence is far from assured. Of the 2,000 survey respondents, only 23% knew Stats NZ reasonably or very well, and only one third viewed the agency positively or very positively. Of those who knew at least a little about Stats NZ or had done the census in the last 5 years ($n = 1,938$), less than half had moderate or high trust in Stats NZ to:

- use personal information to benefit society (47%);
- keep personal information safe (49%);
- be open and transparent in how personal information is managed and used (42%); and,
- take account of public views in decision making relating to personal data (36%, Nielsen 2018).

The number of Māori respondents was too small ($n = 301$) to permit disaggregated results but research carried out by others strongly suggests that knowledge of and trust in Stats NZ would be significantly lower among Māori (Tuhono Trust 2017; Gulliver et al. 2018).

The controversial 2018 Census brings the issue of social licence into sharp relief. To plug major data holes in the census, Stats NZ has drawn extensively on individuals' information from other government datasets. In short, it has devised a method to locate missing individuals in other government data in order to add them into the census dataset (for more details, see Stats NZ 2019b, 2019c, 2019d). This method differs dramatically from previous census practice and bears little resemblance to the understanding of the census process that most New Zealanders had going into Census 2018. The 'administrative enumeration' add-ons equate to 11 per cent of the total approximately 4.7 million people in the census dataset (Stats NZ 2019c).

In a previous paper, we estimated that up to 30% of Maori were missed by the Census 2018 enumeration (Kukutai and Cormack 2018). A recent report by Stats NZ confirmed that the interim collection response rate[2] from individual census forms was just 68.2% for Maori nationally and 65.1% for Pacific peoples (Stats NZ 2019d). Both are far below the interim national response rate of 83.3%. In addition, of the total number of (ethnic) Maori counted in the census dataset, 71% were sourced from individual census forms, 6% were from 'partial responses' (for which ethnic affiliation still had to be sourced elsewhere, see fn 3) and 23% from other government data (i.e., individuals' 2013 census record, birth registration,

Ministry of Education tertiary enrolment etc.). In other words, the use of other government data to plug census holes for Māori far exceeded the 11% for the total population.

In addition to the IDI data, information provided directly from the Ministry of Defence and Department of Corrections was also used to complete the census dataset. This too is a major departure from previous census practice whereby prisoner data were only sourced through prisoners' census forms. According to a Stats NZ report, 'The administrative data that the Ministry of Defence and Department of Corrections provided was a list of everyone in defence establishments or prisons across New Zealand on census night' (Stats NZ 2019b). These records were compared to the information collected through the 2018 Census forms to establish who had already been included in the census data, and who had not. Individuals who were missing from the census were then added in using the records provided directly by the institutions[3].

There is an obvious tension here in that the people who have been disproportionately subject to the use of data undertaken on the basis of purported social licence are the least likely to have a high degree of trust in Stats NZ. In the case of Māori, this trust may have been further eroded by Stats NZ's announcement in April that it cannot provide official iwi data from the 2018 Census due to data quality issues (Stats NZ 2019c). The implications for iwi of this failure are grave given that the census is the only official dataset that reliably collects iwi affiliation data (Kukutai and Cormack 2018). While one might argue that the lack of collective action over Census 2018 reflects Māori (and New Zealanders') tacit acceptance of Stats NZ's social licence, social scientists have good reason to be sceptical. The more likely explanation is that most people simply do not know how their data are being used for secondary purposes in this new integrated data environment.

Seeing like a state

Census 2018 illustrates how marginalised groups that have been the 'subjects' of data 'surveillance' activities (Tufekci 2014) can be disproportionately impacted by data integration activities, but also lack a meaningful mechanism to have their views included in decision-making that affects them (O'Neil 2016; Eubanks 2018). In his influential book S*eeing like a state*, social scientist James Scott traces the ways in which governments have attempted to impose schemas on their populations that make them more legible (Scott 1998). However, these schemas often fail to recognise complex forms of local knowledge and can work directly against the populace's interests. In a big data context, Taylor and Broeders (2015) propose that this 'seeing' also occurs through the power that the state gains from amassing large amounts of data about people, whom may lack an awareness of how their data are being collected, integrated and used. There are multiple ways for 'data misspeak' to occur, especially when the focus is on identifying sub-populations for the purpose of intervening. Data points are not self-evident facts but reflect the social, political and cultural contexts in which the data are collected, analysed and interpreted (boyd and Crawford 2012). In Aotearoa NZ, changes in the data ecosystem means the capability for the government to 'see like a state' is now more profound than ever.

In this regard we see continuities with historical data practices (re)produced as part of colonialism. Data have long been used in colonial nation-states to count, classify, monitor, and construct accounts of Indigenous peoples (Ittmann et al. 2010). Couldry and Mejias (2019) argue that the concept of 'data colonialism' provides a useful frame for thinking about the current big data environment ' … not as a mere metaphor, nor as an echo or simple continuation of historic forms of territorial colonialism, but to refer to a new form of colonialism distinctive of the twenty-first century' (p. 337).

Current data practices also tend to draw on the same extractive logics as those that underpin colonialism and capitalism, especially in relation to natural resources (Couldry and Mejias 2019). These extractive logics also permeate the NZ data landscape, reflected in taglines such as

Stats NZ's declared vision to 'unleash the power of data to change lives'[4] and 'unlocking the economic and social value of data across New Zealand's data ecosystem' (New Zealand Data Futures Forum n.d.). Given that inequities in the distribution of power and privilege have long animated the social sciences, obvious questions arise about who is doing the unleashing and unlocking, on whom, for what purpose, and according to whose values? Drawing on the work of decolonialisation scholar Aníbal Quijano, Couldry and Mejias (2019) identify the fundamental problem with data colonialism as ' … its vision of totality' that excludes other ways of understanding data relations. A re-visioning of data relations and practices is, therefore, essential to interrupt data colonialism.

The potential of Indigenous and Māori data sovereignty

Māori data sovereignty (MD-Sov), and Indigenous data sovereignty (ID-Sov), provide an alternative vision of data practices that are grounded in Indigenous ways of being and knowing and self-determining aspirations (Kukutai and Taylor 2016; Carroll Rainie et al. 2017). MD-Sov refers to the inherent rights and interests that Māori have in relation to the collection, ownership, and application of Māori data (digital or digitisable information or knowledge that is about or from Māori people, language, culture, resources or environments) (Cannataci 2018; Te Mana Raraunga 2018; UN Special Rapporteur on the Right to Privacy 2019). ID-Sov is supported by Indigenous peoples' inherent rights of self-determination and governance over their peoples, country (including lands, waters and sky) and resources as described in the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). Implicit in MD-Sov is the desire for data to be used in ways that support and enhance the collective wellbeing and self-determination of Indigenous peoples – a sentiment emphasised by Indigenous NGOs, communities and tribes (FNIGC 2016; Hudson et al. 2016). In practice ID-Sov means that Indigenous peoples need to be the decision-makers around how data about them are used or deployed.

In contrast to current dominant data practices, MD-Sov grounds data practices within relational ontologies that recognise the inherent dignity and humanity of all peoples. In this sense it provides a framework to transform relationships with data in ways that not only protect against harm for Māori, but also provide transformative data practices for a greater good. MD-Sov, alongside ID-Sov and movements for data justice, prompt us to broaden our focus on the technological possibilities or challenges of datafication, to explore the deeper questions about the types of data relations being (re)produced, how these are shaping broader social relations, and in whose interests they are operating (Dencik et al. 2019). MD-Sov and ID-Sov also challenge taken-for-granted concepts that are grounded in Anglo-centric worldviews such as the primacy of individual privacy.

MD-Sov principles developed by Te Mana Raraunga, the Maori Data Sovereignty Network, provide a useful starting point for operationalising what transformative data practices for collective benefit might look like (Te Mana Raraunga 2018, see Appendix A). Framed within the six Māori values of rangatiratanga, whakapapa, whanaungatanga, kotahitanga, manaakitanga and kaitiakitanga, the 16 principles provide guidance on how to engage with issues such as data control, jurisdiction, disaggregation, accountabilities, guardianship and consent. Returning to the 2018 Census example, principles relating to control and consent are particularly relevant, especially as they relate to decision-making and the significant use of administrative data for Māori. Thus, Principle 5.2 states that 'Free, prior and informed consent (FPIC) shall underpin the collection and use of all data from or about Māori. Less defined types of consent shall be balanced by stronger governance arrangements'. Principle 1.1 states that 'Māori have an inherent right to exercise control over Māori data and Māori data ecosystems'. Clearly, a MD-Sov principle-based approach to Census 2018 would produce a very different process

than that undertaken to date. At the very least, it would require a mechanism for robust data governance. Last year Stats NZ committed to a Treaty-based Māori approach to data governance across the official data system as a way forward (Stats NZ n.d.), but it is unclear how far that has progressed.

It is also useful to consider open data initiatives in Aotearoa NZ given that ID-Sov has 'challenged dominant discourses in open data, questioning current approaches to data ownership, licensing, and use' (Rainie et al. 2019). If we consider the open government data programme[5], approaches could be informed by principles relating to restrictions (6.3 Restrictions: 'Māori shall decide which Māori data shall be controlled (tapu) or open (noa) access)', and guardianship (6.1: 'Māori data shall be stored and transferred in such a way that it enables and reinforces the capacity of Māori to exercise kaitiakitanga over Māori data').

MD-Sov principles challenge the universality of data practices and relations that have become naturalised in our societies, calling attention to their partial and contingent nature. Rather than promoting a framework that can simply be layered over an established Westernised data system, MD-Sov (re)imagines data relations that are transformed, decolonised, and indigenised. We have focused here on data harms and protective mechanisms but also acknowledge the tremendous potential of data to generate individual and collective benefit (see, for example, MD-Sov principle 4.1). The MD-Sov approach has the potential to benefit all groups, particularly those most marginalised within state systems of data surveillance.

The need to attend to these issues is becoming more pressing given the expansion of well-being measurement approaches requiring increasingly granular data, a more permissive inter-agency data sharing environment and well-documented problems of algorithmic bias (O'Neil 2016; Eubanks 2018). The boundaries of what are considered acceptable data collection and use also continue to be stretched. Examples include Stats NZ's 'population density' pilot that uses anonymised aggregated data from telco companies to develop commercial products that track real time movements of people within cities and regions (Moir 2019), and recent media reports of the potential commercialisation of the IDI by charging for a 'value-add service'.[6] Social science in Aotearoa NZ has much to contribute to a critically informed understanding of how processes of datafication are reshaping human relationships and societal structures and institutions, and to hold those in power to account.

Notes

1. The IDI application process is currently being revised to include some Treaty and tikanga considerations.
2. This rate is based on the number of individual forms received, divided by the number of people who should have been counted (using a dual system estimation benchmark). Stats NZ also computed an alternative 'new' response rate which includes partial information (at least two or more of name, date of birth, meshblock) from an individual, dwelling or household summary form. These rates are not comparable with previously computed rates. The interim Maori response rate using this new method was 74.3% and 73.5% for Pacific peoples. The final response rates will be published once the results from the Post Enumeration Survey have been processed in early 2020.
3. The report notes: "Administrative enumerations that were not already in the census were sourced from the records within these datasets". A subsequent report by Stats NZ (2019d) confirms that, "Information on ethnicity has also been provided by the Ministry of Defence and Department of Corrections".
4. See: https://www.stats.govt.nz/about-us/
5. https://www.data.govt.nz/about/open-data-nz/
6. https://www.stuff.co.nz/national/politics/114331277/inside-stats-nzs-gaping-money-hole

References

Boston J, Gill D, editors. 2017. Social investment. A New Zealand policy experiment. Wellington: Bridget Williams Books.

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

Cannataci J. 2018. Report of the Special Rapporteur on the right to privacy. 17 October, A/73/45712. https://www.ohchr.org.

Carroll Rainie S, Schultz J, Briggs E, Palmanteer-Holder P. 2017. Data as a strategic resource: self-determination, governance and the data challenge for Indigenous nations in the United States. The International Indigenous Policy Journal. 8(2):1–29.

Couldry N, Mejias U. 2019. Data colonialism: rethinking big data's relation to the contemporary subject. Television & New Media. 20(4):336–349.

Couldry N, Yu J. 2018. Deconstructing datafication's brave new world. New Media & Society. 20(12):4473–4491.

Dencik L, Hintz A, Redden J, Treré E. 2019. Exploring data justice: conceptions, applications and directions. Information, Communication & Society. 22(7):873–881. doi:10.1080/1369118X.2019.1606268.

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

First Nations Information Governance Centre. 2016. Pathways to First Nations' data and information sovereignty. In: Kukutai T, Taylor J, editors. Indigenous data sovereignty: toward an agenda. Canberra: ANU Press; p. 139–156.

Gulliver P, Jonas M, Fanslow J, McIntosh T, Waayer D. 2018. Surveys, social licence and the integrated data infrastructure. Aotearoa New Zealand Social Work. 30(3):67–71.

Hudson M, Farrar D, McLean L. 2016. Tribal data sovereignty: Whakatōhea rights and interests. In: Kukutai T, Taylor J, editors. Indigenous data sovereignty: toward an agenda. Canberra: ANU Press; p. 157–178.

Ittmann K, Cordell D, Maddox G, editors. 2010. The demographics of empire: colonial order and the creation of knowledge. Athens, OH: Ohio University Press.

Kukutai T, Cormack D. 2018. Census 2018 and implications for Maori. New Zealand Population Review. 44:131–151.

Kukutai T, Taylor J, editors. 2016. Indigenous data sovereignty: toward an agenda. Canberra: ANU Press.

Mann M, Daly A. 2019. (Big) data and the North-in-South: Australia's informational imperialism and digital colonialism. Television & New Media. 20(4):379–395.

Mittelstadt B, Floridi L. 2016. The ethics of big data: current and foreseeable issues in biomedical context. Science and Engineering Ethics. 22:303–341.

Moir J. 2019, 19 June. Stats NZ, phone companies to track people's movements. Radio New Zealand.

New Zealand Data Futures Forum. n.d. Harnessing the economic and social power of data.

Nielsen. 2018. A social licence approach to trust. A close-up on trust. https://www.stats.govt.nz/corporate/measuring-stats-nzs-social-licence.

O'Neil C. 2016. Weapons of maths destruction. How big data increases inequality and threatens democracy. London: Penguin.

Rainie S, Kukutai T, Walter M, et al. 2019. Issues in open data – imdigenous data sovereignty. In: T. Davies, S. Walker, M. Rubinstein, F. Perini, editor. The state of open data: histories and horizons. Cape Town and Ottawa: African Minds and International Development Research Centre. https://www.stateofopendata.od4d.net/chapters/issues/indigenous-data.html.

Scott J. 1998. Seeing like a state. How certain schemes to improve the human condition have failed. New Haven: Yale University Press.

Stats NZ. 2018. Algorithm assessment report. https://data.govt.nz/use-data/analyse-data/government-algorithm-transparency.

Stats NZ. 2019a. Integrated Data Infrastructure. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/.

Stats NZ. 2019b. Creating the 2018 Census dataset by combining administrative data and census forms data: Our privacy impact assessment. https://www.stats.govt.nz/privacy-impact-assessments/creating-the-2018-census-dataset-by-combining-administrative-data-and-census-forms-data-our-privacy-impact-assessment.

Stats NZ. 2019c. 2018 Census – real data about real people. https://www.stats.govt.nz/news/2018-census-real-data-about-real-people.

Stats NZ. 2019d. 2018 Census: Interim coverage rates, collection response rates, and data sources. https://www.stats.govt.nz/reports/2018-census-interim-coverage-rates-collection-response-rates-and-data-sources.

Stats NZ. n.d. Co-designing Māori data governance.

Taylor L, Broeders D. 2015. In the name of development: power, profit and the datafication of the global South. Geoforum. 64:229–237.

Te Mana Raraunga. 2018. *Maori data sovereignty principles*. https://www.temanararaunga.maori.nz/new-page-2.

Tufekci Z. 2014. Engineering the public: big data, surveillance and computational politics. First Monday. 19(7). https://firstmonday.org/article/view/4901/4097.

Tuhono Trust. 2017. *Sharing information for wellbeing. Maori engagement on social license report*. http://datafutures.co.nz/our-work-2/talking-to-new-zealanders/.

UN Special Rapporteur on the Right to Privacy. 2019. *Draft recommendation on the protection and use of health-related data*. https://www.ohchr.org/EN/Issues/Privacy/SR/Pages/HealthRelatedData.aspx.

van Dijck J. 2014. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. Surveillance & Society. 12:197–208.

Appendix A: Māori Data Sovereignty Principles

01 Rangatiratanga | Authority

1.1 Control. Māori have an inherent right to exercise control over Māori data and Māori data ecosystems. This right includes, but is not limited to, the creation, collection, access, analysis, interpretation, management, security, dissemination, use and reuse of Māori data.

1.2 Jurisdiction. Decisions about the physical and virtual storage of Māori data shall enhance control for current and future generations. Whenever possible, Māori data shall be stored in Aotearoa New Zealand.

1.3 Self-determination. Māori have the right to data that is relevant and empowers sustainable self-determination and effective self-governance.

02 Whakapapa | Relationships

2.1 Context. All data has a whakapapa (genealogy). Accurate metadata should, at minimum, provide information about the provenance of the data, the purpose(s) for its collection, the context of its collection, and the parties involved.

2.2 Data disaggregation. The ability to disaggregate Māori data increases its relevance for Māori communities and iwi. Māori data shall be collected and coded using categories that prioritise Māori needs and aspirations.

2.3 Future use. Current decision-making over data can have long-term consequences, good and bad, for future generations of Māori. A key goal of Māori data governance should be to protect against future harm.

03 Whanaungatanga | Obligations

3.1 Balancing rights. Individuals' rights (including privacy rights), risks and benefits in relation to data need to be balanced with those of the groups of which they are a part. In some contexts, collective Māori rights will prevail over those of individuals.

3.2 Accountabilities. Individuals and organisations responsible for the creation, collection, analysis, management, access, security or dissemination of Māori data are accountable to the communities, groups and individuals from whom the data derive.

04 Kotahitanga | Collective benefit

4.1 Benefit. Data ecosystems shall be designed and function in ways that enable Māori to derive individual and collective benefit.

4.2 Build capacity. Māori Data Sovereignty requires the development of a Māori workforce to enable the creation, collection, management, security, governance and application of data.

4.3 Connect. Connections between Māori and other Indigenous peoples shall be supported to enable the sharing of strategies, resources and ideas in relation to data, and the attainment of common goals.

05 Manaakitanga | Reciprocity

5.1 Respect. The collection, use and interpretation of data shall uphold the dignity of Māori communities, groups and individuals. Data analysis that stigmatises or blames Māori can result in collective and individual harm and should be actively avoided.

5.2 Consent. Free, prior and informed consent (FPIC)[1] shall underpin the collection and use of all data from or about Māori. Less defined types of consent shall be balanced by stronger governance arrangements.

06 Kaitiakitanga | Guardianship

6.1 Guardianship. Māori data shall be stored and transferred in such a way that it enables and reinforces the capacity of Māori to exercise kaitiakitanga over Māori data.

6.2 Ethics. Tikanga, kawa (protocols) and mātauranga (knowledge) shall underpin the protection, access and use of Māori data.

6.3 Restrictions. Māori shall decide which Māori data shall be controlled (tapu) or open (noa) access.

---

1 https://www.un.org/development/desa/indigenouspeoples/ publications/2016/10/free-prior-and-informed-consent-an-indi   genous-peoples-right-and-a-good-practice-for-local-communities-fao/

Tahu Kukutai
National Institute of Demographic and Economic Analysis, The University of Waikato,
Hamilton, New Zealand
✉ tahu.kukutai@waikato.ac.nz

Donna Cormack
Te Kupenga Hauora Māori, University of Auckland, Auckland, New Zealand
ⓘ http://orcid.org/0000-0003-2854-3595