

Surveys, social licence and the Integrated Data Infrastructure

Pauline Gulliver, Monique Jonas, Tracey McIntosh, Janet Fanslow and Debbie Waayer,

University of Auckland, New Zealand

ABSTRACT

INTRODUCTION: Statistics New Zealand's Integrated Data Infrastructure (IDI) is a central repository for researchers to access multiple government agency datasets. The aim of this investigation was to understand social licence for including survey data in the IDI.

METHODS: Two convenience samples were recruited: (1) participants in one of 10 focus groups; and (2) respondents to pilot surveys for the 2018 NZ census or a population-based survey on violence experience. Qualitative data were transcribed and analysed using thematic analysis. Analyses were conducted independently by two members of the research team and results compared.

FINDINGS: Whilst little prior awareness of the IDI existed, participants developed considered judgements about it, identifying concerns and proposing safeguards that would encourage them to support its maintenance and use.

CONCLUSIONS: While there is the potential for social licence to be granted for the IDI, an on-going, transparent engagement process is required to maintain trust with agencies and researchers. As an over-represented population within government agency data, active, honest engagement is required with Māori, as are safeguards to reduce risks of further stigmatisation and marginalisation.

KEYWORDS: big data; social licence; indigenous data; policy development

Introduction

"[The] quest to understand and explain what works and for whom in what circumstances underpins the notion of evidence-based policy making" (Sanderson, 2002). In 2014, Statistics New Zealand (StatsNZ, Aotearoa New Zealand's (NZ's) independent agency for the collection and delivery of robust, independent statistics (Statistics New Zealand, 2017c)) launched the Integrated Data Infrastructure (IDI), which "contains person-centred data from a range of government agencies, Statistics NZ surveys, and non-government organisations"

(see Figure 1). The IDI provides a "one-stop-shop" for researchers to access multiple government datasets to understand health, justice, education, social services and income outcomes. While linkage is undertaken at the individual level, analysis occurs at the aggregate level by academic and policy development researchers (Statistics New Zealand, 2017a). The creation of linked datasets for research purposes is not new (Holman et al., 2008). However, the difference between bespoke linkages to answer specific research questions and large-scale, linked datasets developed as a repository lies in the use, re-use, and

AOTEAROA
NEW ZEALAND SOCIAL
WORK 30(3), 57–71.

CORRESPONDENCE TO:
Pauline Gulliver
p.gulliver@auckland.ac.nz

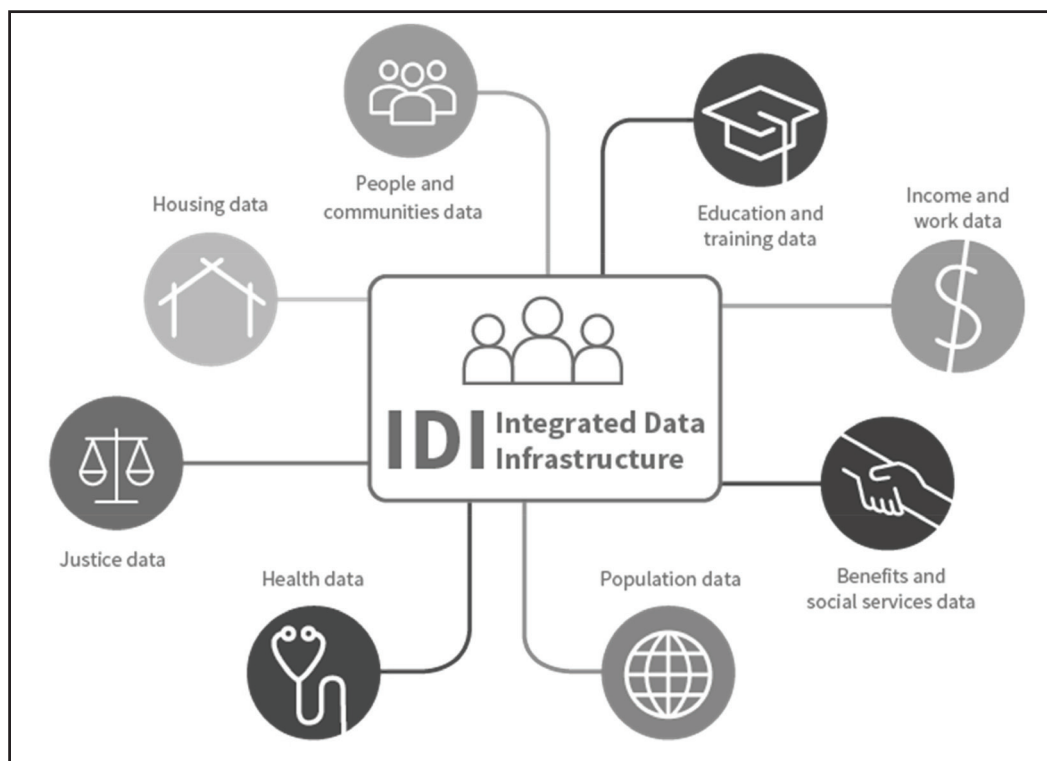


Figure 1. Overview of the IDI, from http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx, accessed 25 September 2017.

potential unknown future use of such data sets.

To date, the IDI has largely eluded the NZ public's consciousness. However, it is increasingly used as evidence for policy decisions, such as restructuring funding mechanisms for state-funded schools (Edwards, 2016). Governmental support for the IDI has resulted in pressure on non-governmental organisations and researchers in receipt of government funding to make client databases available for inclusion in the IDI as a condition of their funding (Kirk, 2016). Since the change in government following the 2017 general election, the discourse around the use of large-scale, linked data sets has subtly changed: "while the numbers are critical, the insight gained from doing the analysis is just as important" (Social Investment Agency, 2018). Further, the NZ government has launched an investigation into how "government currently uses algorithms, to give New

Zealanders confidence that their data is being used appropriately" (Curran & Shaw, 2018).

The principles and regulatory mechanisms governing the establishment and use of large-scale, linked datasets such as the IDI are subject to increasing debate (Casanovas, De Koker, Meddelson, & Watts, 2017). In NZ, four Health and Disability Ethics Committees (HDECs) review health and disability research. However, secondary analysis of de-identified administrative data for observational studies is exempt from HDEC review (Health and Disability Ethics Committee, 2014). Because personally identifying information within the IDI is withheld or encrypted, analyses utilising IDI data are considered exempt. While some users of the IDI, particularly university researchers, are required to obtain ethics approval through university ethics committees, StatsNZ does not require approval from an ethical review committee

prior to granting access to the IDI, or peer review to ensure research rigour.

There are, however, safety mechanisms in place for use of the IDI:

- Referee checks are conducted on researchers;
- Only data necessary to answer research questions are made available;
- Access to the data is through a controlled data environment;
- Only research questions considered to have a wider public interest are considered.
- Output from research conducted using IDI data is checked to ensure results are not personally identifiable (Statistics New Zealand, 2017a).

International experiences reveal challenges for broad-scale data linkage and use. Developing and employing large datasets containing microdata can engender public distrust (Presser, Hruskova, Rowbottom, & Kancir, 2015), and political anxiety about invasion of privacy (Dudley-Nicholson, 2016). Recent English experience illustrates how a failure to secure public trust can fatally jeopardise efforts to harness the potential of 'big data'. Care.data was a central repository of health and social care data from all English National Health Service funded care settings, developed primarily for research and evaluation purposes. Established in 2013, its implementation became so problematic that it was discontinued in 2014. Contributing issues included inadequate management and communication, a contradictory legislative environment with respect to patient confidentiality (Presser et al., 2015), the intention for Care.data to be accessible to private companies, and an all-or-nothing opt-out consent process (Shaw, 2014). Its collapse was attributed to failure to attend to levels of public trust in health services and research, and the conditions upon which that trust is based, combined with scepticism about Care.data's public good orientation. The concept of social licence has

been employed to explain the failure of Care.data, (Carter, Laurie, & Dixon-Woods, 2014). It is also invoked by StatsNZ as a necessary condition of the success of the IDI.

The aim of the current investigation was to elucidate the extent of social licence for including survey data in the IDI. As with related concepts such as trust, social licence is not straightforward to delineate or measure (Boutilier & Thomson, 2011). Both the concept of trust and the requirements of social licence are likely to: mean different things to different people (Rooney, Leach, & Ashworth, 2014); vary by context (Hall, Lacey, Carr-Cornish, & Dowd, 2015) and be highly sensitive to changing public perceptions (Heikkinen, Lepy, Sarkki, & Komu, 2016). Despite its ineffability, experience drawn from a range of fields suggests that social licence can be a decisive force in allowing or preventing a range of activities (Rooney et al., 2014). Therefore, it is useful to understand social licence as it pertains to the IDI and what conditions or parameters members of the public might place upon it, whilst acknowledging its dynamic, variable nature.

Social licence

The concept of social licence derives from the work of sociologist Everett Hughes (Hughes, 1958), 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. This permission constitutes social licence – it implies a mandate, empowering the licensed agent to ask things of others in relation to the licensed practice (Hughes, 1958, p. 78).

Social licence has risen to prominence particularly within industries and enterprises that impose harms upon resources or communities, such as mining, forestry and fishing. A "social licence to operate" is predicated upon an agent meeting "the expectations of society regarding the conduct and activities of corporations that go beyond

the requirements of formal regulation” (Carter et al., 2014), p. 404). It emerges from a process. Trust is central in enabling and sustaining social licence, representing a willingness to accept vulnerability to the actions of another in some domain (Mollering, 2006; Rousseau, Sitkin, Burt, & Camerer, 1998). We were interested to learn how people conceived themselves and others as vulnerable, and what determined willingness to accept vulnerability (and therefore trust the process) and confer social licence.

We adopted the following working definition of social licence:

Societal acceptance that a practice that lies outside general norms may be performed by a certain agent, on certain terms. It is the result of an ongoing process of negotiating terms with a wider societal group, and means that the practice can be performed by that agent without incurring social sanction. Social licence confers a mandate upon the licensee to ask things of others in relation to the licensed practice.

Our definition makes explicit that the practice under consideration lies outside of general norms. Specifically, the linkage of individual-level information collected by government agencies without explicit knowledge of the people from whom data were collected. It differs from the definition adopted by the Data Futures Partnership which limited social licence to “When people trust that their data will be used as they have agreed, and accepted that enough value will be created, they are likely to be comfortable with its use” (Data Futures Partnership, 2016).

It is important to note at the outset that this work is focussed specifically on social licence as it relates to the linkage of data for analysis at the aggregate level for research and policy development purposes. This differs from the sharing of information for the purposes of service delivery, a process regularly

undertaken, often with the explicit consent from the individual and understanding of the process involved.

Methods

As social licence cannot be conferred if the relevant community is unaware of the agency seeking it, the research team were interested in investigating whether members of the NZ public were aware that the IDI existed; were able to understand the practice lying outside social norms involved in using the IDI; and the terms considered necessary to engage in this practice without incurring social sanction. Rather than attempting to quantify the overall level of social licence, we investigated awareness of, and attitudes towards, the IDI amongst two separately selected convenience samples. One comprised respondents to one of two surveys with different levels of perceived sensitivity; the other were participants in focus groups.

Respondents to the survey were “primed” about the nature of the survey by answering

Table 1. Demographic Characteristics of the Interview Sample (Frequencies)

	Census	Sensitive
Age		
15-24 yrs	2	9
25-34 yrs	12	7
35-44 yrs	4	4
45+ yrs	13	11
Sex		
Male	13	11
Female	17	20
Gender diverse	1	1
Ethnicity		
NZ European	25	14
Māori	0	8
Other	6	10
Total	31	32

a series of questions, administered face-to-face by an interviewer. The goal was to determine if the nature of the survey was associated with the likelihood of consenting to linkage. Survey 1 was a subset of census questions administered by StatsNZ interviewers ($n = 31$ participants); Survey 2 was a subset of questions designed to assess violence exposure in the general population ($n = 32$ participants). Age, sex and ethnicity of survey participants are presented in Table 1. There is an important distinction in the recruitment methods employed by StatsNZ and the university. The university research team was particularly interested in obtaining feedback from sections of the population who are either over-represented in violence statistics, or for whom very little information about the prevalence of violence is reported. As such, recruitment methods were designed to encourage participation from Māori, LGBTIQ+ and disabled members of the community.

Following the survey questions, participants were asked open-ended questions about:

- whether they would consent to have the information they had just provided uploaded and linked with government agency data;
- their reasons to provide or withhold consent;
- what safeguards would favourably influence their decision; and
- what information about process (where would the data be stored) and usage (access rights, long-term storage) would assist an informed choice.

Semi-structured focus groups were conducted to contribute to the design of a population-based survey on violence exposure. Participants were asked about:

- Their understanding of the information NZ government agencies hold about individuals and what is done with it.
- The degree to which they considered it acceptable to link survey data with information held by government agencies:

- in general;
 - if they were data participants had personally provided (to the survey);
 - if the data were anonymised;
 - if the focus of the research was at the population level, rather than focusing on individuals;
 - if the information they provided could be used again in the future for unknown purposes.
- Their views on the safeguards that should be in place for the data to be made available for other researchers.
 - Who should store the combined administrative and survey-based data.
 - Whether there are types of data that they felt could be shared without seeking individual permission, or types of data that it would be wrong to seek permission to share.

The focus groups comprised the following community groups:

- Disabled persons;
- Members of the LGBTIQ+ community;
- Mothers of young children;
- New migrants;
- Young Māori men;
- Māori women (×2);
- Pākehā men;
- Intimate partner violence (IPV) survivor advocates;
- Māori (mixed gender);
- Mixed (general).

Interview data were recorded with an electronic data capture programme, REDCap (P. A. Harris et al., 2009) and hand-written notes. Focus group discussions were digitally recorded and transcribed. All data were analysed using thematic analysis, an inductive method of analysis which explores the manifest (content that is noted or mentioned directly by respondents) and latent (implicit or underlying) themes. Analyses were conducted independently by two members of the research team and results compared. Where interpretation differed, the analyses were brought to the wider research team and resolved through discussion.

Results from interviews and focus group discussions were analysed and are presented separately.

Ethics approval for this investigation was granted by the University of Auckland Human Participants Ethics Committee (ref 017300). All participants provided informed consent prior to participating.

Results

Arm 1: Interviews with Survey Respondents.

Most (81% of census and 70% of violence survey) survey respondents indicated they would be willing to have their completed survey linked with government agency data. Where participants did not consent to linkage, they were more likely to have disclosed a social norm breach. Violence survey participants were asked about their pornography use – 30% of those who reported pornography use would consent compared with 63% of those who reported no pornography use.

Amongst those willing to consent to data linkage there was a perception that this would “help the government make better decisions” (Survey respondent 1 [SR1]) and “provides access to more comprehensive data” [SR6]. Respondents to both surveys also indicated their data were “already out there” [SR23], alluding to information

collected through social media, or that they “didn’t really have anything to hide” [SR18]. Those who would not consent were concerned with how their information would be used, who would access it, and whether they would be identified:

Not right to link up my story from separate sources. [SR32]

No knowledge of who may access it. Concern around confidentiality, inappropriate use, lack of clarity around destruction dates. [SR24]

A total of 23% of the violence survey and 45% of census survey participants wanted additional information to help them decide whether they would consent to having their survey data linked with government agency data, including who would access it, what it would be used for, and what protections were in place:

Who can access government agency data? What will government use it for? [SR3]

Explain the purpose and process of data sharing in the IDI and a guarantee around privacy. [SR16]

Survey participants were provided with a range of options that might influence their willingness to provide consent to have their information linked. The proportion of participants from each survey who

Table 2. What Would Make it Easier to Consent to Have Your Information Linked?

Option	Census Sample	Sensitive Sample
Assurance that my name and address will be removed from the data I provide	74%	81%
Guarantee that the information is only available to bona fide researchers	65%	65%
Freedom to be able to withdraw the data whenever I want	58%	58%
Knowledge that the information was held by Statistics New Zealand	58%	50%
Assurance that the information would be destroyed after a set period of time	52%	46%
Knowledge that the information was held by a university	39%	46%

agreed to each option (more than one could be selected) are summarised in Table 2. Across the two groups, there was consistent agreement with the importance of knowing that their name and address would be removed, control over access, and being able to withdraw the data.

Arm 2: Focus Groups:

All focus group participants were aware that government agencies collected information about them. There was acceptance of information sharing for the purposes of service provision.

...you collect it once then lots of people can, if it's safe, use the same information instead of asking again. [Mother of young children]

...changing your GP they transfer the data from one place to another and they have full information on health issues and everything. [New migrant]

However, few participants were aware of the IDI. Extended discussion was needed to shift thinking from individual-level data sharing for service delivery, to population-level data linkage for service and policy design.

Five key themes emerged from discussions: (1) good quality data are important; (2) understanding the context of data collection; (3) privacy is important; (4) oversight of the researchers is required; and (5) yes, I am happy, if I provide consent.

1. Good quality data are important

Views concerning the quality of the data available for data linkage could be connected to participant experiences with individual agencies. For example, IPV survivor advocates highlighted concerns about the quality of the information collected:

I have seen hundreds of [Statutory Social Service Agency] files from women that have asked the ministry... what's written

about them and honestly... none of it based on fact. [IPV Survivor]

Members of the disabled community highlighted problems associated with communicating and how this could impact on the quality of information collected and recorded about them.

It gives me chills really when I go to anywhere like a GP cos there's such a fear of misdiagnosis for deaf people because they just don't understand us. It's 5 to 10 percent of deaf that get their information accurately passed on so I think there's a big fear in the deaf community about the representation of us. [Disabled person]

The pervading concern expressed by these groups was that the IDI could ultimately use incorrectly recorded data. This would be amplified for community members with a higher degree of interaction with government agencies, who subsequently have more information collected about them.

2. Understanding the context of data collection

A key concern for Māori was understanding the context of data and collection and understanding the meaning of data. These concerns were mirrored by representatives of marginalised populations as well:

...data that [are] given, in ... an interview can be de-contextualised and alienated from what the [original] kaupapa [purpose, rationale] was. [Māori mixed gender]

[the data] becomes kind of powerless because all the information...is completely different...so [can] easily...get distorted...out of context. [LGBTIQ+]

Even if data were factually correct, their meaning could be lost or distorted if the context of collection was not understood. Linked to this were concerns that structural and societal context could be obscured by the focus on individual-level data. An important

manifestation of this relates to Māori over-representation in government datasets that report negative social indicators, a situation reflecting the harms of colonisation, societal disparities and institutional racism (Kukutai, 2011) which may not be recognised and properly accounted for by those who use the data. These risks were highlighted by participants who could imagine the IDI being used to compound damaging stereotypes, especially for communities that had more information held about them:

...it's going to highlight our problems...
It is often used just to prove cultural deficits. [Māori women]

I come from a kura kaupapa [Māori language immersion school] and we just have different aspirations and I think that is what we should be measured by. We come from Māori aspirations, from our whānau collective not from a government collective. But in saying that, we didn't [our kura] under-achieve, we over-achieved for our region, but that is because it came from our own people. However we are not making our overall data look better because we are a small proportion. Māori students in mainstream schools are still [expected] to live up to the aspirations of government. Their data [are] collected by Pākehā as well. [Māori mixed gender]

3. Privacy is important

Privacy concerns, nominated by all participants, were largely expressed in terms of the risk of individual identification:

[T]he risk comes when it becomes personal and [discloses] personal information that might be scrutinized by a government department for their purposes. [LBGTIQ+]

Some participants commented on the difficulties of ensuring non-identifiability when collating many pieces of personal information:

As soon as you put my age, my impairment, the length of time I've had my impairment – there are some people who would know exactly who I am.
[Disabled person]

What information was perceived as requiring greater protection differed between individuals and groups. Privacy concerns were raised about financial information (Pākehā men); social services and police records (Māori, IPV survivor advocates); and health service information (Disabled persons and mothers of young children).

For some, individual control over information was connected with trust in political processes and players:

Coming from a country where control over your information is very highly valued and sharing that information has sometimes had really detrimental impacts on people I think just as an individual, knowing where my information is going to go, and having some control over that is important. [New migrant]

...our demographics are changing...a lot of the people that are coming into NZ, I'm not saying that they are corrupt, but they come from corrupt countries where corruptness is a way of life. So in 30 or 50 years' time, how is this information going to be treated? [Pākehā man]

4. Oversight of the researchers is required

Participants identified a need for research accountability – an independent body that should have oversight of the research process.

...as long as safeguards were in place and appropriate processes were in place to make sure that information is only accessed and used in a certain way, that has to be really clear. [New migrant]

The most solid, robust oversight process is what people would feel most comfortable

with because there are so many people who will want access to that data because data is so terribly valuable...people who have some teeth, who can say no and can remain unchallenged. [Pākehā man]

5. Yes, I am happy to have my data linked, with my consent

Despite their concerns, the majority of participants would be willing to allow researchers to use the data stored about them within the IDI if they were asked. Trust was an enabler of consent, as was the perception that data linkage would benefit themselves, their family, or the community. Participants with negative experiences with government agencies were less likely to approve. Participants viewed the consent process as both required and expected.

I know that that's my information, I know...that that's my record...I just want to be asked. [Māori women]

When prompted and provided with the opportunity, focus group participants were interested in and willing to engage with the concept of the IDI. They could see the benefits as well as risks and highlighted opportunities for strengthening the securities around access:

...as far as being able to pull information out for research purposes, I think it's amazing to have all that data linked and to know you've got the same individual across data systems; and there's just so much you can do with all that information. [New Migrant]

While clear that consent was necessary, participants were realistic about the challenges resulting from a consenting process:

Isn't that where the greater good comes in? Governments, whether you like it or not have to have information on the populous of the country...you can't opt in and out of that sort of thing, it's just not practical. [Pākehā man]

Discussion

The input from our participants provides insights into the nature of the vulnerabilities those living in New Zealand perceive in relation to the IDI, along with the conditions under which they may be willing to make themselves vulnerable and place trust in those analysing and using the data for the sake of the good that they anticipate may be derived from use of the IDI. This informs an understanding of the conditions that may be placed upon the IDI's social licence. Whilst little prior awareness of the IDI existed amongst our participants, they developed considered judgements about it, identifying concerns and proposing safeguards that would encourage them to support its maintenance and use. Fairly high levels of institutional trust in the integrity and competence of StatsNZ and related agencies present in our sample were tempered with suspicion, often borne from experience. This reinforces the view that social licence is an on-going process of negotiation, dependent on judgements over an agent's integrity and competence (Butler & Cantrell, 1984; Mayer, Davis, & Schoorman, 1995).

The expressed willingness to grant social licence, whilst strong, was not unconditional: guarantees are needed about data quality; researcher awareness of, and sensitivity to, the structural and societal context surrounding data; and that individuals will not be identifiable. Robust oversight mechanisms are required to ensure appropriate use of this taonga (treasure / gift). Further, Māori expressed concern about how decisions around data use were to be made. This speaks to the notions of data ownership and data sovereignty.

A strength of the research presented is the methods used to derive an understanding of social licence for linking data in different contexts. When interview participants were provided with the opportunity to report a social norm breach (use of pornography) in confidence, and then asked about whether they would consent to having that data

linked with government agency data, they were less likely to do so. These findings suggests there are limits to acceptable sharing of data. Where those data have the potential to be damaging in the future, it appears that individuals are less likely to trust the process. Indeed, those who expressed fewer concerns about sharing data indicated they had “nothing to hide”, implying no perceived breaches of social norms. The focus group discussions allowed a discussion amongst peers about these limits. For example, the men’s and young mother’s focus groups moved from indicating they had “nothing to hide” to discussing concerns around the use of information they *did* consider sensitive (financial data and health disorders, respectively).

1. Good quality data are important

Government agency administrative data form the basis of the IDI. As administrative data vary in quality (Daas, Ossen, & Tennekes, 2010), problems with the raw data collected from service-level interactions is subsequently embedded in the IDI, and carried over into future use. Participants with significant experience of government agencies were readily able to identify this risk.

To date, the IDI has been used as the basis for two predictive risk models (for child maltreatment, and educational non-achievement) that, if implemented, could influence a child’s life trajectory. Although developed at the population level, implementation of predictive models at the individual level has been contemplated. Therefore, despite the application of StatsNZ’s safety mechanisms that prevent the *publication* of results that could be personally identifiable, it is possible that the *application* of research derived from the IDI could impact at the individual level. It is well established that population-level risk measures seldom translate well to the individual level (Rockhill, Kawachi, & Colditz, 2000). The failings of population-

level risk prediction based on suboptimal data will be greater still. Examples of incomplete or inconsistent reporting are easily identified in NZ (Gulliver, Cryer, & Langley, 2013; Mansell, Ota, Erasmus, & Marks, 2011) as they are in the majority of developed countries (Jansen, 2012). Because of the non-random nature of missing data in the majority of social measures, imputation is not appropriate where data are absent (Sterne et al., 2009). The drive for the use of linked datasets for policy development needs to be balanced with investment to ensure good quality data are collected, recorded, and transferred appropriately. Without this, there is a risk that both competence-based and integrity-based elements of public trust will be eroded, and social licence correspondingly compromised. Our findings demonstrate the permeable interface between trust of government agencies and the IDI itself: participants’ views about the competence and integrity of agencies they had dealings with informed their views about the IDI. Securing social licence for the IDI requires attention to data integrity at the provider agency level, as well as within the IDI.

2. Context of data collection is important

The promise of the IDI is that previously elusive associations will be revealed through data linkage. However, to understand and interpret identified associations, the original context and purpose of data collection must be understood. Participants who had experienced unfair assumptions by agencies were particularly likely to identify this need. In NZ, Māori are over-represented in the majority of adverse outcomes (McIntosh, 2011; McIntosh & Coster, 2017), and are more likely to come into contact with government agencies, increasing the data collected about them. Understanding the drivers for over-representation is important (Cram, Gulliver, Wilson, & Ota, 2015). Institutional racism, enacted in government policies, and interpersonal racism, enacted through discriminatory behaviour, is demonstrated to be negatively associated with health

outcomes for Māori (R. Harris et al., 2006) and Indigenous and minority populations elsewhere (Jones, 2001). Indeed, some key determinants of health outcomes (ethnicity and gender) are socially defined constructs which influence life experiences (Jones, 2001).

In countries like NZ, whose colonial history impacts upon current health and social outcomes, power within the research space is an important contextual issue. The Treaty of Waitangi establishes the responsibility of the Crown to uphold the rights of Māori. Te Mana Rauanga (the Māori Data Sovereignty Network) have identified three areas of focus to uphold Treaty obligations with respect to the IDI. These include social licence, the expectation that the government will act in the interests of Māori; cultural licence, the impact of data integration and sharing on the social contract that exists through the Treaty; and Māori data sovereignty, recognising that Māori data should be subject to Māori governance (Hudson, 2016).

Māori are often not included in research design even where research focuses upon them, or is particularly salient to Māori. Utilising deficit-focussed data is likely to reproduce deficit perspectives and further entrench power differentials (Walter, 2016). Recently, StatsNZ has opened its first international data lab, providing access to the IDI for researchers outside NZ (Statistics New Zealand, 2017b). This underscores the need to ensure that researchers understand the NZ cultural context.

3. Privacy is important

Privacy is central to research ethics, and is established in NZ law through the Privacy Act (1993). Participant concern that data should not enable individuals to be identified is reflected in Principle Ten of the Privacy Act ("Privacy Act," 1993), which states that information obtained by an agency for a given purpose should not be used for another purpose, unless one

of seven permitted exceptions applies, including that the information: "is used in a form in which the individual concerned is not identified" (10.(f) (i)); or "is used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned" (10.(f) (ii)). Principle Three states that, where personal information is collected from a person, the agency must take reasonable steps to ensure that the person is aware that information is being collected (3.(1)(a)); the purposes for which it is being collected (3.(1)(b)), and the intended recipients of the information (3.(1)(c)).

While StatsNZ apply safe reporting practices to protect privacy (Statistics New Zealand, 2017a), the potential for re-identification is real (Malin, Karp, & Scheuermann, 2015). Attention to identifiability reflects the role that privacy is seen to play in protecting individuals from adverse consequences. Such concerns might be indicated in the apparent association between disclosed social norm breach (pornography use) and lack of willingness to consent to survey linkage revealed in our findings. This suggests that ensuring information in the IDI cannot be linked back to individuals will be crucial in securing and maintaining social licence for including survey data in the IDI. However reasons to protect privacy extend beyond adverse consequences for the individual into respect for persons (Benn, 1984; Fried, 1968); preserving liberty (Hallborg, 1986); and delimiting state power (Solove, 2007). The preference our participants expressed for IDI data use to be predicated upon consent suggests that their concerns about privacy extend beyond direct adverse consequences into the social norms establishing the terms of engagement between state and individual. Privacy is often seen as an umbrella concept with implications for how information is collected, stored, used, disseminated and applied (Solove 2007). Whilst research conducted using the IDI has been at the population level, some proposed applications are at the individual level

(Edwards, 2016). The concern about privacy returns the focus to the strong links between trust in the institutions using the data and granting social licence for the IDI. Concerns expressed that such information could once again be used to discriminate and reinforce prejudices are based on historical realities rather than abstract paranoia.

4. Who watches the researchers?

Oversight from an independent, trusted agency, was identified by some participants as a way to minimise potential harms. This finding resonates with experience from mining, in which the existence of a trusted agency with decision-making rights has been seen to support social licence (Prno, 2013). One response to the risks our participants identified would be to require ethics committee approval to access the IDI. Training for committee members in the potential harms and benefits of statistical analysis of large de-identified population datasets would be necessary to ensure that the review process provided the intended protections. However, this measure cannot guarantee protection if the data collected were already prejudiced. Māori have critiqued university ethics procedures as “Eurocentric”, privileging liberal notions of the “autonomous individual participant” rather than considering collectivist constructs to guide the research process, resulting in a “condescending ethos” (Tuari, 2014, p.134). Potential responses include developing:

...a Māori-dominated ethics process that is dedicated to supporting Māori post-graduates, established researchers and non-Indigenous scholars wanting our guidance on conducting ethical research with Māori. A process of this kind will focus in part on holding [research ethics boards], government agencies and private research accountable if their conduct negatively impacts Māori researchers and *Māori research participants* (Tuari, 2014, p. 146, emphasis added).

Other options include citizen juries and public engagement campaigns (van Staa, Goldacre, Buchan, & Smeeth, 2016). Van Staa and colleagues have called for accountability, transparent operations, and data stewardship including staff training, standard operating procedures and a focus on the skills and attitudes of staff to ensure appropriate data usage (van Staa et al., 2016).

In a systematic review of research into public acceptability of data linkage, Aitken, de St Jorre, Pagliari, Jepson, and Cunningham-Burley (2016) acknowledged the importance of consent as a consistent theme of public responses to the linkage of health data. However, the authors also highlighted that the need for consent appeared to be strongly associated with trust in the institutions, organisations or individuals involved in processing or accessing their data:

...rather than focussing on which consent mechanisms are most favoured by members of the public, it may be more valuable to focus on how relationships of trust are built up (and conversely eroded) and how trust can be facilitated within research and data-sharing or data-linkage processes including through public / patient engagement or involvement. (Aitken et al., 2016, p. 15)

Once again, this highlights the transient nature of social licence for data linkage if agencies and researchers do not behave as good stewards of the taonga.

Transparency and public engagement appear to have been at the forefront of data linkage initiatives in Scotland, drawing on lessons from the English Care.data experience, and seeking to maintain social licence. The Farr Institute of Health Informatics Research holds four regional public panels that scrutinise and advise on governance systems, public engagement plans and research practices, as well as additional virtual panels, forums and public panels (Farr Institute). The Scottish Primary Care Information

Resource has an independent advisory group that reviews requests for the use of data, as well as an opt-out process (NHS National Services Scotland, 2016). The advisory group includes patient representatives, general practitioners and specialist confidentiality advisors.

5. Yes, I am happy, with my consent

Many participants expected consent to be sought for data use. This was perceived as important even when participants were willing to consent. This may reflect the idea that respect for persons requires acknowledgement of their rights and interests, and consequently permission for activities that may impact upon them (Benn 1984). The same ideas underpin social licence processes and align strongly with recognition of the significance of Indigenous Data Sovereignty. Data Sovereignty is linked with Indigenous Peoples' rights to:

maintain, control, protect, and develop their cultural heritage, traditional knowledge and traditional cultural expressions, as well as their right to maintain, control, protect and develop their intellectual property over these... If Indigenous Peoples have control over what and how data and knowledge will be generated, analysed and documented, and over the dissemination and use of these, positive results can come about. (Kukutai & Taylor, 2016, p. xxii)

The consent condition may also reflect a concern for transparency as a check upon agency power, a mark of respect for citizens and a way of protecting public good orientation (Taylor, 2011). The expectation for individual-level consent is a challenge to projects like the IDI, and one that requires serious reflection from agencies such as StatsNZ. Whilst our participants expressed an expectation to consent, they also appreciated the logistical barriers. Whilst individual consent for use of all data is not feasible, governance structures involving lay members and ensuring representation

of special interest groups are feasible and embody respect for those whose information is contained within the IDI, and whose interests may be affected by its use. In support of Te Mana Rauranga (Hudson, 2016), we argue that the Treaty of Waitangi, which guarantees Māori control over taonga (treasures), requires strong Māori input into how the IDI is used.

Conclusion

As Rooney et al. (2014, p. 211) have observed, "in many vitally important respects a [social licence] is constituted by knowledge and meaning, rather than by legal documents and permits instituted through a bureaucratic-administrative mechanism." Participants in this study appreciated the purpose and potential of the IDI when it was explained to them, but they lacked pre-existing knowledge of it. They identified concerns and suggested safeguards that would reassure them sufficiently to consent to inclusion of their data within the IDI. We conclude that, while there is the potential for social licence to be granted for the IDI, an on-going, transparent engagement process is also required that provides individuals with the ability to interact with research and policy initiatives being developed in this space. As an over-represented population within government agency data, active, honest engagement is required with Māori, and safeguards to reduce the risks of further stigmatisation and marginalisation are required.

Acknowledgement

The authors would like to greatly acknowledge the support of the Data Futures Partnership for providing partnership funding for this research, and Statistics New Zealand for accommodating the research in the build-up to census 2018. Further, we would like to acknowledge the constructive feedback provided by anonymous peer reviewers for the journal.

References

- Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R., & Cunningham-Burley, S. (2016). Public responses to the sharing and linkage of health data for research purposes: A systematic review and thematic synthesis of qualitative studies. *BMC Medical Ethics, 17*(73). doi:10.1186/s12910-016-0153-x
- Benn, S. I. (1984). Privacy freedom and respect for persons. In F. Shoemann (Ed.), *Philosophical dimensions of privacy: An anthology* (pp. 223–244). Cambridge, UK: Cambridge University Press.
- Boutilier, R., & Thomson, I. (2011). Measuring and modelling the social licence to operate: Fruits of a dialogue between theory and practice. Retrieved from <https://sociallicense.com/publications.html>
- Butler, J., & Cantrell, R. (1984). A behavioural decision theory approach to modelling dyadic trust in superiors and subordinates. *Psychological Reports, 55*, 19–28.
- Carter, P., Laurie, G. T., & Dixon-Woods, M. (2014). The social licence for research: Why care.data ran into trouble. *Journal of Medical Ethics, 41*(5), 404–409. doi:10.1136/medethics-2014-102374
- Casanovas, P., De Koker, L., Medderson, D., & Watts, D. (2017). Regulation of big data: Perspectives on strategy, policy, law and privacy. *Health and Technology, 1*–15. doi:10.1007/s12553-017-0190-6
- Cram, F., Gulliver, P., Wilson, M., & Ota, R. (2015). Understanding over-representation of indigenous children in child welfare data: An application of the Drake risk and bias models. *Child Maltreatment, 20*(3), 170–182.
- Curran, C., & Shaw, J. (2018). Government to undertake urgent algorithm stocktake [Press release]. Retrieved from <https://www.beehive.govt.nz/release/government-undertake-urgent-algorithm-stocktake>
- Daas, P. J. H., Ossen, S. J. L., & Tennekes, M. (2010). *Determination of administrative data quality: Recent results and new developments*. Retrieved from Heerlen, The Netherlands http://pietdaas.nl/beta/pubs/pubs/Q2010_Session34_paper.pdf
- Data Futures Partnership. (2016). *Exploring social licence*. Retrieved from datafutures.co.nz/assets/Uploads/DFP-Engagement-doc-FINAL.pdf
- Dudley-Nicholson, J. (2016, August 11). Census 2016: Australia has become a global joke. *The Daily Telegraph*. Retrieved from <http://www.dailytelegraph.com.au/rendezview/census-2016-australia-has-become-a-global-joke/news-story/0491d8ffd72da480da27483c39e5043b>,
- Edwards, D. (2016). *Review of finding systems for early learning and schooling: Predictive modelling of risk of underachievement*. Retrieved from <https://education.govt.nz/assets/Documents/Ministry/Information-releases/All-for-release.pdf>
- Farr Institute. (n.d.) *Public panels*. Retrieved from <http://www.farrinstitute.org/public-engagement-involvement/public-panels>
- Fried, C. (1968). Privacy. *Yale Law Review, 77*, 475–493.
- Gulliver, P., Cryer, C., & Langley, J. (2013). Monitoring trends in violence: A delayed response to Estrada (2006). *Journal of Interpersonal Violence, 28*(16), 3129–3148.
- Hall, N., Lacey, J., Carr-Cornish, S., & Dowd, A. (2015). Social licence to operate: Understanding how a concept has been translated into practice in energy industries. *Journal of Cleaner Production, 86*, 301–310.
- Hallborg, R. (1986). Principles of liberty and the right to privacy. *Law and Philosophy, 5*, 175–218.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, J., & Conde, J. G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381.
- Harris, R., Tobias, M., Jeffreys, M., Waldegrave, K., Karlsen, S., & Nazroo, J. (2006). Effects of self-reported racial discrimination and deprivation on Maori health and inequalities in New Zealand: Cross-sectional study. *The Lancet, 367*(9527), 2005–2009.
- Health and Disability Ethics Committee. (2014). *Standard operating procedures for health and disability ethics committees*. Retrieved from <http://www.ethics.health.govt.nz>
- Heikkinen, H., Lepy, E., Sarkki, S., & Komu, T. (2016). Challenges in acquiring a social licence to mine in the globalising Arctic. *Polar Record, 52*(265), 399–411.
- Holman, C. D. J., Bass, J. A., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J., . . . Stanley, F. J. (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.
- Hudson, M. (2016). Social licence: A Maori's perspective. Retrieved from <https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/591304ea414fb570382a3c1e/1494418668414/Social+license.pdf>, 18 June 2018
- Hughes, E. C. (1958). *Men and their work*. Glencoe, IL: Free Press.
- Jansen, H. A. F. M. (2012). *September Prevalence surveys on violence against women. Challenges around indicators, data collection and use*. Paper presented at the Expert Group Meeting: Prevention of violence against women and girls, Bangkok, Thailand.
- Jones, C. P. (2001). Invited commentary: “Race”, racism, and the practice of epidemiology. *American Journal of Epidemiology, 154*(4), 299–304.
- Kirk, S. (2016, 18 October). Government cuts 5000 children from Growing Up in NZ longitudinal study. *Stuff*. Retrieved from <http://www.stuff.co.nz/science/85464263/Government-cuts-5000-children-from-Growing-up-in-NZ-longitudinal-study>,
- Kukutai, T. (2011). Contemporary issues in Maori demography. In T. McIntosh & M. Mulholland (Eds.), *Maori and social issues* (pp. 11–48). Wellington, NZ: Huia.
- Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous data sovereignty. Toward an agenda*. Canberra, ACT: Australian National University.
- Malin, B., Karp, D., & Scheuermann, R. H. (2015). Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. *Journal of Investigative Medicine, 58*(1). doi:10.2310/JIM.0b013e3181c9b2ea

- Mansell, J., Ota, R., Erasmus, R., & Marks, K. (2011). Reframing child protection: A response to a constant crisis of confidence in child protection. *Children and Youth Services Review, 33*(11), 2076–2086.
- Mayer, R., Davis, J., & Schoorman, F. (1995). An integrative model of organisational trust. *Academic Management Review, 20*(3), 709–734.
- McIntosh, T. K. (2011). Maori and cross-cultural research: Criticality, ethicality and generosity [Special issue]. *New Zealand Sociology, 26*, 61–75.
- McIntosh, T. K., & Coster, S. (2017). Indigenous insider knowledge and prison identity. *Counterfutures, 3*, 69–99.
- Mollering, G. (2006). *Trust: Reason, routine, reflexivity*. Oxford, UK: Elsevier Press.
- NHS National Services Scotland. (2016). *SPIRE: Your questions answered*. Retrieved from <http://spire.scot/your-questions-answered/>
- Presser, L., Hruskova, M., Rowbottom, H., & Kancir, J. (2015). Care.data and access to UK health records: Patient privacy and public trust. *Technology Science, 2015081103*, 1–35.
- Privacy Act, New Zealand Government, Pub. L. No. 28 (1993 1 July 1993).
- Prno, J. (2013). An analysis of factors leading to the establishment of a social licence to operate in the mining industry. *Resources Policy, 38*, 577–590.
- Rockhill, B., Kawachi, I., & Colditz, G. A. (2000). Individual risk prediction and population-wide disease prevention. *Epidemiological Reviews, 22*(1), 176–180.
- Rooney, D., Leach, J., & Ashworth, P. (2014). Doing the social in social licence. *Social Epistemology, 28*(3–4), 209–218.
- Rousseau, D., Sitkin, S., Burt, R., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academic Management Review, 23*(3), 393–404.
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration, 80*(1), 1–22.
- Shaw, D. (2014). Care.data, consent and confidentiality. *Lancet, 383*(April 5), 1205.
- Social Investment Agency. (2018). How do we invest for social wellbeing? Retrieved from <https://sia.govt.nz/about-us/what-is-social-investment/>, 13 June 2018
- Solove, D. (2007). I've got nothing to hide and other misunderstandings of privacy. *San Diego Law Review, 44*, 745–772.
- Statistics New Zealand. (2017a). How we keep IDI data safe. Retrieved from http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx#safes
- Statistics New Zealand. (2017b). International data lab approved in Sydney. Retrieved from http://www.stats.govt.nz/tools_and_services/media-centre/media-releases-2017/stats-opens-int-data-lab.aspx
- Statistics New Zealand. (2017c). Who we are. Retrieved from http://www.stats.govt.nz/about_us/who-we-are.aspx
- Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., . . . Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal, 338*, b2393. doi:<https://doi.org/10.1136/bmj.b2393>
- Taylor, M. C. (2011). Health research, data protection and the public interest in notification. *Medical Law Review, 19*, 267–303.
- Tuari, J. (2014). Resisting condescending research ethics in Aotearoa New Zealand. *AlterNative: A Journal of Indigenous Peoples, 10*(2), 134–150.
- van Staa, T.-P., Goldacre, B., Buchan, I., & Smeeth, L. (2016). Big health data: The need to earn public trust. *British Medical Journal, 354*. doi:[10.1136/bmj.i3636](https://doi.org/10.1136/bmj.i3636)
- Walter, M. (2016). Data politics and Indigenous representation in Australian statistics. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda*. Canberra, ACT: Australian National University.