Surveys, social licence and the IDI

A collaborative project between
The University of Auckland

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

Statistics New Zealand

Gayle Beck, Andrea Lawson, Matthew Flanagan

Co-funded by the Data Futures Partnership

7 Dec 2016
Structure of workshop

1. Background and introduction
2. Social licence
3. Results of interviews and focus groups
4. Results from Maori participants
5. Workshop
6. Summaries and wrap-up
Why undertake this research?

- There is a strong push from central government for the inclusion of population-based surveys into the integrated infrastructure (IDI)

- The practicalities of inclusion raise anxiety levels for academic researchers due to:
  - The need to hand over data
  - Uncertainty over future use
  - Concern over impact on response rates
The importance of ethics committees in university research

The University of Auckland Human Participant Ethics Policy:

“No project involving human participants can be carried out by staff or students of The University of Auckland without the approval of The University of Auckland Human Participants Ethics Committee”

The University of Auckland Code of Conduct in Research:

“The Education Act 1989 protects the academic freedom of academic staff and students to undertake research, but this academic freedom is predicated on the need to maintain the highest ethical standards”
Key ethics requirements

(i) **Autonomy**  Participants should freely consent to their participation in the research

(ii) **Beneficence**  acting in the public good; it includes all actions which are intended to promote the good of other people.

(iii) **Non-maleficence**  research should minimise and manage risks of harm, such as the risk of physical or psychological harm

(iv) **Justice**  researchers have a duty to ensure that the benefits of their research are achieved through just means; that the benefits and burdens of research are fairly distributed; and that there is fair treatment in the recruitment of participants.
Informed consent and data custodianship

“the researcher must provide participants with adequate information about the purpose of the research, methods of participant involvement, and intended use of the results”

“Data stored for the purpose of the original research should be accessible by the researcher and supervisor only...Storage of data for posterity and future research that involves transfer to a public repository requires a suitable release form negotiated with the participant that clarifies conditions of future access”
Surveys and social licence

Data

Consent

Ethics
Social Licence

A conceptual Analysis

Wednesday, 7 December 2016
Integrated Data Infrastructure (IDI)
Structure

1. What is Social Licence and When is it Needed?
2. What Norms Are Infringed by the IDI?
3. Mandate
4. How Does Social Licence Work: 2 examples
5. Lessons and Implications for the IDI
What is Social Licence and When Is It Needed?

- Norms confer permissions and prohibitions
- Prohibited activities are liable to sanction
- Social licence needed to perform prohibited activities without sanction.

‘A professional has a licence to deviate from lay conduct…it is an institutionalized deviation, in which there is a certain strain towards definition of situation and roles.’(Hughes 1963:656)
Conceptualisations of Social Licence

“A social licence to operate refers to the ongoing acceptance and approval of a mining development by local community members and other stakeholders that can affect its profitability’ (Moffat and Zhang 2014:61).

“When people trust that their data will be used as they have agreed, and accept that enough value will be created, they are likely to be more comfortable with its use. This acceptance is referred to as social licence.’ (Data Futures Partnership 2016:3)
What Social Licence Is and Is Not!

NOT:
A tool for pacifying dissenters

IS:
Genuine and needed authorisation from wider group - who have the authority to give or withhold licence
Social Licence: A proposed working definition

‘Social licence is 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 a process of negotiation with a wider societal group, and means that the practice can be performed by that agent without incurring social sanction.’
What norms are infringed by the IDI?

- Privacy
- Respect for Autonomy (control over projects to which one contributes)
a power, implied by the licence, for the agent:

‘to define what is proper conduct of others towards the matters concerned with their work’ (Hughes 1958:78).

• Terms of the licence and the mandate are open to negotiation.

• What is the content of the licence and mandate sought for the IDI? (What powers would it give, and what would it demand of the public?)
Revised Working Definition:

‘Social licence is 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 a process of negotiation 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 licencsee to ask things of others in relation to the licensed practice.’
How Does Social Licence Work?

Example 1: Driving

Obvious physical threat

Legislated early:

Low initial tolerance of risk gradually increased

Macro and micro aspects to licence (authorities and drivers)
Example 2: Medicine

- Physical and social/reputational risks
- Hypocratic oath: early recognition of social licence?
- Unregulated practitioners & dubious benefits led to public suspicion
- Professionalisation and regulation have led to more trust
- Terms of social licence tightened after scandals
Corfield on the growth of trust in apocatharies:

3 main drivers:
- Strong demand for treatment
- Demonstration of public good
- Increasing professionalization

‘The medical professions wielded influence over their patients, but the practitioners were also themselves regulated. It was and remains an intensely socialised process, based upon negotiation and trust. Such an outcome was only possible after a long pre-history of micro-change in the reputation and practices of eighteenth-century medicine.’ (Corfield 2009:17)
Implications/Lessons for IDI:

- Social Licence takes time to earn
- Cautious beginnings can lead to greater latitude
- Demonstrable social benefit supports social licence
- Maintaining social licence requires ongoing responsiveness
- Social licence has macro and micro dimensions
- Formal regulation can assist, but does not guarantee, social licence
- Social licence is ambiguous and transitory
- IDI could inherit trust and/or distrust from associated agencies
Project aims and our approach
Aims

• To develop an understanding of the added securities necessary for linking “sensitive” as opposed to “neutral” data.

• Consider recommendations for the development of appropriate and trustworthy data management systems for linking data into the future.
Our approach

Exploratory:
• Semi-structured questions following an interview that “primed” respondents about the type of questions asked.
  • Predominantly face-to-face; some online (n=12)

• Focus groups, discussion points included
  ➢ Understanding of “Public Good Research”
  ➢ Understanding of the types of data collected by government agencies
  ➢ Factors that would influence the likelihood of consenting to have survey data linked with government agency data

Ethics approval provided by the UoA Human Participants Ethics Committee
Participants

Census respondents

• Convenience sample recruited through poster ads, facebook ads, key contacts

Sensitive survey respondents

• Recruited to represent key “at risk” communities.

• Convenience sample recruited through:
  • Email
  • Reference groups
  • Key contacts
# Interview participant characteristics

<table>
<thead>
<tr>
<th></th>
<th>Census</th>
<th>Sensitive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-24 yrs</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>25-34 yrs</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>35-44 yrs</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>45+ yrs</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Female</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>Gender diverse</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NZ</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>European</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Maori</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>31</td>
<td>32</td>
</tr>
</tbody>
</table>
Focus groups

- Disabled persons
- LGBTIQ+
- Mothers of young children
- New Migrants
- Young people
- Young Maori Men
- Maori Women (x2)
- Men
- Survivor advocates
- Maori (mixed)
- Mixed (general)

- Transcribed
- Thematically analysed
- Consensus approach to analysis
Results from interviews and focus groups
Quick reminder of what we did

63 x short individual interviews
- To understand whether survey context influences consent to link
- Asked post exposure to either a ‘sensitive’ or ‘less sensitive’ survey

12 x focus groups
- An in-depth understanding of participants’ social license to link survey data with government agency data
Definitions

“Shared” data
Data shared within and across government agencies, often at an individual level for a service delivery outcome

“Linked” data
Data from different government agencies linked together in a database for a research/policy outcome
Data **sharing** is **expected** and **accepted** for the purposes of **service provision**

- Assumption that data is currently shared
- Potential for greater sharing than currently exists
- Sharing must be for personal benefit

“I am hugely in favour of this – it is hugely invaluable”  
[Man]

“It depends if it is just between practitioners and you agreed that they send it over is one thing but if they start to share it with say your employment or boss or something”  
[New migrant]

“I have heard stories of people having difficulty accessing things from the doctor that they needed, they haven’t been able to share information that I thought you could”  
[Mum]
There is a lack of understanding about data linking

- General misunderstanding on difference between linked data and shared data
- When explained many could see the value of a linked dataset
- There were common concerns about who would have access and for what/whose benefit

“I think it’s really important, they have to have the information from somewhere. Policy development – it’s important they get it right”  
[mums]

“Decision makers don’t always act positively on the good information”

[Men]

“I’m not sure what you would do with it, maybe it could identify someone that would be in a situation of [ ], and then do what, a knock on the door, find it hard to imagine how it would work”

[Survivor]
Limited awareness & mixed response to IDI

- Most participants were unaware of the existence of IDI
- Some can rationalise the benefit as long as sufficient safeguards are in place
- Some were alarmed: data quality and anonymity were key concerns
Participants may be less likely to consent to linking data they consider particularly sensitive

- There is no set definition of what constitutes ‘sensitive’ data
- Most participants would consent to link ‘sensitive’ data if sufficient safeguards in place
- Participants that provided information they considered particularly ‘sensitive’ may be less likely to consent to linking
- Awareness of linkage may impact respondent disclosure, thus impacting data quality

“I have heard parents say you don’t want to get involved with the mental health system because if that information is shared there is a stigma attached, so that is going to negatively influence outcomes in other areas of their life”

[MUM]

“I think the more people think that bugger is creeping into your personal life the less they are going to disclose”

[LGBTIQ+]
Most participants would consent to having their data linked….

Please note: for illustrative purposes only, not based on quantitative data
Participants want informed consent

• Permission should be obtained for all data linkage
• Sufficient safeguards need to be in place and communicated
• The data linkage must be for a beneficial purpose (personal or public)

“if you are doing things for my improvement, or for my betterment, then it’s fine”

[New Migrant]

“This is for the benefit of New Zealand…I would not have a problem with it at all”

[Mum]

“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]
Data linkage must be for a beneficial purpose

<table>
<thead>
<tr>
<th>Personal Good</th>
<th>Public Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Reduce respondent burden</td>
<td>• Policy shaping</td>
</tr>
<tr>
<td>• Personal data “one stop shop”</td>
<td>• Helping communities</td>
</tr>
<tr>
<td></td>
<td>• Adds richness to analysis</td>
</tr>
<tr>
<td></td>
<td>• Longitudinal research</td>
</tr>
<tr>
<td>“the cost in the sense that you collect it once then lots of people can,</td>
<td>“Basically it should be that it’s part of developing society and making</td>
</tr>
<tr>
<td>if it’s safe, use the information instead of asking again”</td>
<td>society better”</td>
</tr>
<tr>
<td>“its amazing … for service providers to know exactly what’s gone on with a</td>
<td>“the findings may help, may get the numbers to put things in place to help</td>
</tr>
<tr>
<td>person”</td>
<td>families to see how they can be helped”</td>
</tr>
<tr>
<td>[Mum]</td>
<td>[New Migrant]</td>
</tr>
<tr>
<td>[New Migrant]</td>
<td>[Mum]</td>
</tr>
</tbody>
</table>
A few participants would consent by default

- Got nothing to hide, don’t care, don’t have a view, my info is already out there

“I guess I am a bit of a push-over”

[Maori men]

“I don’t know if that is a bit of an inter-generational difference because we do openly share, and we do quite a lot of information out there without thinking about it, you know on social media and that, and we know that people can trace it, but I think, I don’t know I am not really bothered by it, as long as it’s anonymous”

[Men]

“I couldn’t care less as long as you can’t pinpoint me”

[Mixed]
Participants shared common concerns

<table>
<thead>
<tr>
<th>Fear of surveillance</th>
<th>Fear of discrimination</th>
<th>Fear of disclosure of personal info</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Data used <strong>to check for wrong doing</strong>, or to deny services…</td>
<td>• Concern about <strong>profiling analysis</strong></td>
<td>• <strong>How anonymous</strong> can data from minority groups be?</td>
</tr>
<tr>
<td>• Fear of “<strong>big brother</strong>”, of the government having “your whole story”</td>
<td>• <strong>More</strong> data collected on those in <strong>greater need</strong> of government support</td>
<td>• <strong>Sold for commercial gain</strong></td>
</tr>
<tr>
<td>“the risk comes when it becomes personal and disclosing personal information that might be scrutinized by a government department for their purposes” [LBGTIQ+]</td>
<td>“negative stereotypes that get potentially reinforced by the data” [New migrant]</td>
<td>• <strong>How will my data be used in the future?</strong></td>
</tr>
<tr>
<td>“doesn’t target the elite as much as people in low socio-economic areas” [LBGTIQ+]</td>
<td>“It’s really difficult to keep confidential and anonymous with people who have a degree of difference in the community.” [Disability]</td>
<td>• <strong>Hackers/deliberate misuse</strong></td>
</tr>
</tbody>
</table>
Participants shared common concerns

<table>
<thead>
<tr>
<th>Poor data quality</th>
<th>Appropriateness of reusing data</th>
<th>Appropriate use of data by researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>Subjective data</strong> incorrectly recorded &lt;br&gt;• Poor data quality = meaningless results</td>
<td>• Can data collected for one purpose be re-used for another?</td>
<td>• Access to data &lt;br&gt;• How is it used to inform policy?</td>
</tr>
</tbody>
</table>

“I have seen hundreds of CYFS files from woman that have asked the ministry……..and ask them what’s written about them and honestly …mm and none of it based on fact”<br>[Survivor]

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

“we have no way of know how accurate information is that’s being shared; so typically people in the deaf community are being misinterpreted, or not asked or whatever.”
[Disability]
Trust is an enabler of consent

Yes, definitely consent
Yes, consent but with permission
No, definitely not consent
Institutional trust and an optimistic outlook mean consenters expect good outcomes despite being able to identify concerns.

**Trust**

**Institutional trust**

“it is going to come down to trust and you’ve signed something that says you’re not going to do this, and you been in the job this long, and you’ve passed all the checks so presumably you can be trusted with that”

[Mum]

**Optimistic outlook**

“I don’t have anything to hide, so unless someone has bad intentions I don’t know what they can do about it”

[Mum]

**Distrust of institutions**

“If you let go control of your personal information now, you’ll never get it back”

[Men]

**Negative experiences with service providers**

“when you have received a bad response from the service, it doesn't understand the situation, you are very uncomfortable about sharing at all, to anyone”

[Survivor]
The key concerns of the non-consenters are so great that they can see no value in linking data

**Data quality too poor**

Subjective data not accurately recorded, e.g. police, courts, child records

“well the information is still stuffed”

[Survivor]

**Can never be anonymous**

Unusual records will always be identifiable even if data is de-identified.

“As soon as you put my age, my impairment... there are some people who would know exactly who I am”

[Disability]
Negative experiences with government agencies drives distrust and generates a very emotional response

<table>
<thead>
<tr>
<th>Distrust</th>
<th>Lack of control over my story</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don’t trust government:</td>
<td>Its not right to link up “my story” from separate sources – it’s an invasion of privacy</td>
</tr>
<tr>
<td>- to securely hold or make sensible use of the data</td>
<td>“when you’ve got agencies taking that story and sharing it around it just feels like its another thing that you are losing”</td>
</tr>
<tr>
<td>- Not to use the information against people</td>
<td>[Survivor]</td>
</tr>
<tr>
<td>- To change the data to suit its own need</td>
<td></td>
</tr>
</tbody>
</table>
Key take outs

• There is limited understanding of data linkage and limited awareness that the IDI exists
• The research suggests that most people will consent to having their data linked for a beneficial purpose (personal or public), and as long as sufficient safeguards are in place
• Participants would like an informed consent process to linking their data
• Anonymity is expected, but reassurance is needed
• Institutional trust is an enabler of consent
Māori responses
Acknowledgements

- Led by Tracey McIntosh (Ngä Pae o Te Maramatunga)
- Stephanie Palmer
Public good research

Concept closely aligned with “mainstream” health

_I typically associate that with...the idea of majority who we are not really a part of when they talk about things like the mainstream, upper middle class Pākehā, urban dwelling people. That is what I think of._

The level of concern and distrust of data was apparent.

• distrustful about research process.
• concerned how data could be used to further embed negative stereotypes
• creating an environment for over-surveillance
Admin data and policies for Maori

It’s going to highlight our problems

It is often used just to prove cultural deficits.

If we are perceived as a minority and as long as everyone else is trucking along then let it keep going

Any data can be sensitive
De-contextualising data

“data that is given in say an interview can be de-contextualised and alienated from what the kaupapa was”

“you’ll just be another statistic to …. They put you in the category of New Zealand Maori, yeh yeh and you are part of that group yeh, yeh.”

“...my police record, I didn’t quite understand how they got all that information, ...I didn’t know that they held against me what I had done when I was 15 to when I was 19 and had a kid .....so they told me I was a drunk and drug and alcoholic when I was 16...they were still holding that against me when I was 19... if they had asked me first because I would have had time to explain myself that behaviour was when I was 15. I was being silly and stupid and now I am 40 and I want to get a job, you know in between these years if you look through those years and I was really good”
I would hope that it has gone through ethics and stuff and whoever is collecting the data would be bound by those guidelines as well.

Researchers as kaitiaki
it might just be old school yeh, to be honest. Oh can I use your...um glasses please you know, instead of just going and taking them and...come in and oh you’ve got my glasses on didn’t even ask

Given your previous discussions do you think that linking data will get better outcomes for Maori?

At what stage do you define an outcome by?

If things are...changing, new technologies then that should be emphasised. I feel this would inform the way that people choose to participate. Issues of confidentiality and anonymity seem to [be] fading away.
Workshop
(your turn)
Three options

1. People are comfortable with their data being shared for their care. What process do we need to go through to acquire that social licence for the IDI?

2. Take the results literally and seek consent in every case – what are the implications?

3. Consider the sensitivity of data – should we be taking a different approach depending on the sensitivity of the data collected?

Structure:
15 Minutes discussing the topic
Brief presentation back to the group
Workshop discussions

(Notes)
People are comfortable with their data being shared for their care. What process do we need to go through to acquire that social licence for the IDI?

- Establish (making visible?) a transparent governance process
- Public education on the IDI
  - What it can do
  - What it can’t do
  - Difference between sharing and linking
  - Demonstrating the value of the IDI
  - Protections that are currently in place
- Transparency about the policy uses of the data
- Communicating the independence of StatsNZ and the Government Statistician.
Take the results literally and seek consent in every case – what are the implications?

- Consent at collection, i.e. MSD and MoH
  - Include consent statements
  - For health survey, don’t link if don’t consent
  - For admin data “consent” is agreed via a contract for services => opt out is to opt out of

  ***This is not consent (coercion) – could risk isolating more people***

- But, points where we would need individual consent include:
  - Collection
  - Integration
  - Project applications

- Radical options:
  - Enable individuals to opt out of IDI or specific uses?
  - IDI uses a “My Data” type of portal enabling access to and perusal of, control of and options to opt out
  - Seek consent for types of data use (rather than project by project)

- Challenges:
  - Would need to verify identity for safe access
  - Data is probabilistically matched, what if it is not right?
  - IDI is full population coverage – opt out could skew / create bias
  - Who would consent for rights of children? Deceased?

Logistical challenges: Full population, admin data collection. How? Real Me?
Consider the sensitivity of data – should we be taking a different approach depending on the sensitivity of the data collected?

Who decides what data is sensitive? What is the threshold?
   OPC / HRC / Referendum / Govt Statistician / Peers

Process:
• Vetted and approved purpose
• Authorised people
• Some data further restricted
• All projects through an ethics lens
• Transparency – education and debate to lead to informed trust (issues around budgets, channels and interest)
• Don’t try and separate operational vs research data because all data should lead to improved outcomes and follow a similar process.
A guest presenter

Kevin Sweeny
StatsNZ
Data Quality
Steady States (Quality) Criteria

<table>
<thead>
<tr>
<th>Quality Dimensions - Statistics NZ Quality Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Relevance</td>
</tr>
<tr>
<td>• Accuracy</td>
</tr>
<tr>
<td>• Timeliness</td>
</tr>
<tr>
<td>• Accessibility</td>
</tr>
<tr>
<td>• Consistency</td>
</tr>
<tr>
<td>• Interpretability</td>
</tr>
<tr>
<td>• Safety (proposed)</td>
</tr>
</tbody>
</table>

- Increased LOB Granularity and Measures
  - Preparation (fix errors, format, classify, geocode)
  - Data Knowledge (compare results to real world)
  - Expectations Analysis (alignment with what's expected)
  - Sense-checking (do results make sense?)

- Statistical (Business)
  - Discovery & Access/Versioning (metadata, naming conventions)
  - Supporting Information/Outputs (metadata, data dictionary, process, products)
  - Re-Use (metadata, standards)

- Data/Information Management
  - Storage/Preservation (metadata, archiving, disposal)
The wrap-up
Key Messages

• Many people are willing to consent - consent is seen as an important step.

• Groups with negative experiences are less likely to do so.

• Data quality matters

• Trust needs to be built - recording of information must be
  – Relevant
  – Accurate
  – Neutral

• There needs to be safeguards across all stages of the research process.
  – De-identification doesn’t remove all of the risks.
  – Consider consent, storage, access, analysis, reporting
  – Clarity of purpose
  – Context of data collection
Data Futures Principles
Bibliography


