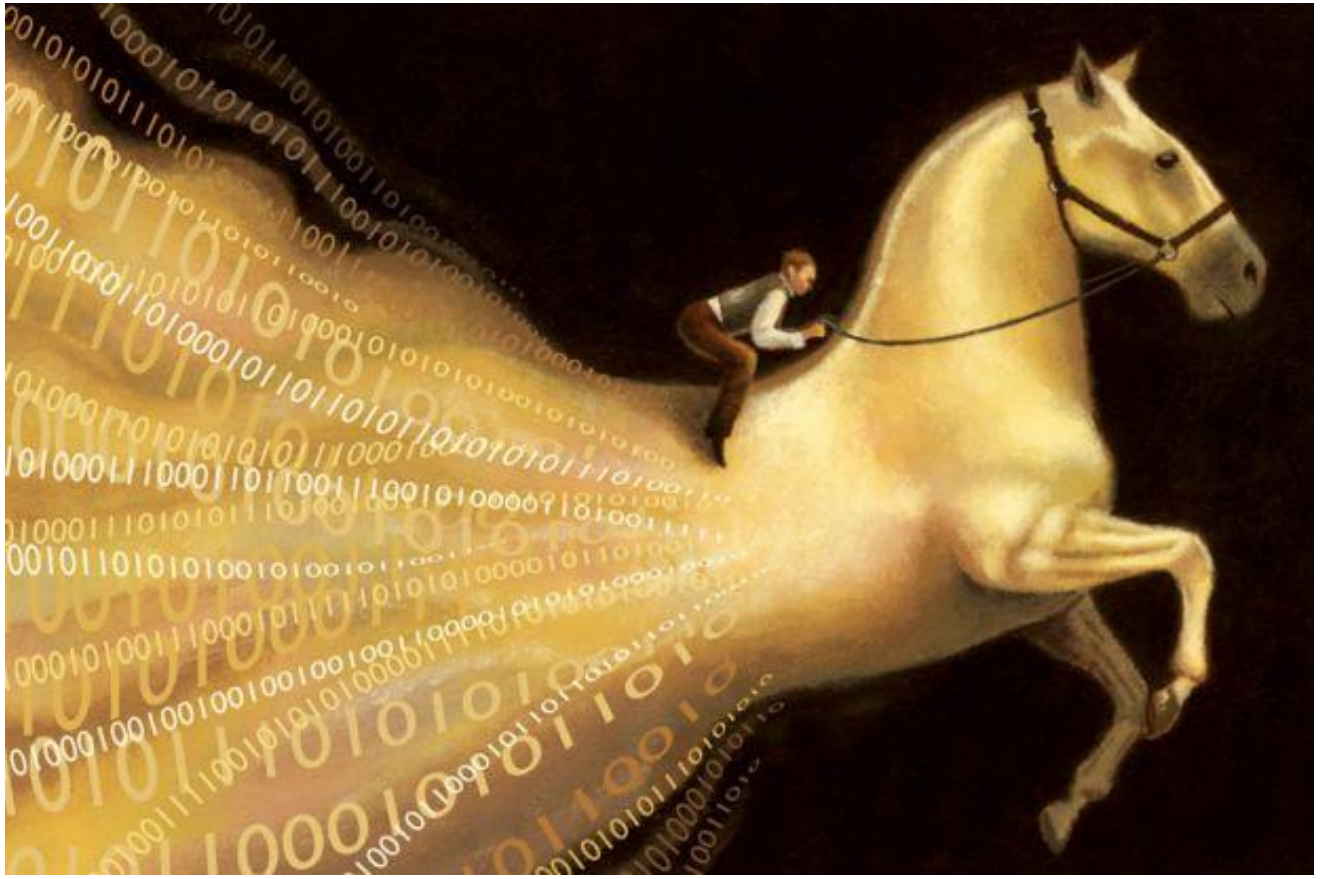
Dr Angela Ballantyne
Angela.ballantyne@otago.ac.nz
PHSS Data Ethics 13.02.2017

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

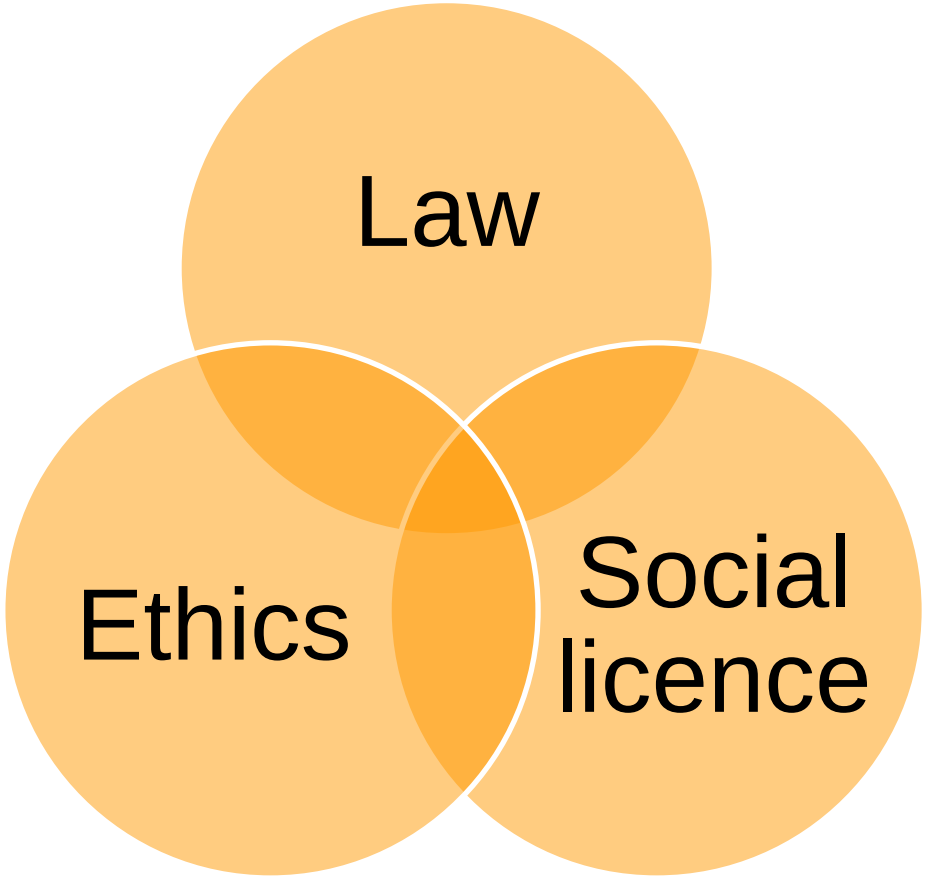
Introduction to data ethics



Overview

- Challenges
  - Demand
  - Sources
  - Using data
- Control
  - Consent
  - Governance
  - Transparency

Law, ethics, social licence



Data

- Surveys, questionnaires
- Clicks, likes, text analysis, links are all data
- What did Google do differently?
- Netzer, Lemaire, Herzenstein study of Prosper, a peer-to-peer lending site (Seth Stephens-Davidowitz 2017)

- God
- promise
- debt-free
- minimum payment
- lower interest rate
- will pay
- graduate
- thank you
- after-tax
- hospital

Data

- Surveys, questionnaires
- Clicks, likes, text analysis, links are all data
- what did Google do differently?
- Netzer, Lemaire, Herzenstein study of Prosper, a peer-to-peer lending site (Seth Stephens-Davidowitz 2017)

- God
- promise
- debt-free
- minimum payment
- lower interest rate
- will pay
- graduate
- thank you
- after-tax
- hospital

Challenge: Increasing Demand

- Improve health service provision
- Reduce costs (social investment)
- Generate new medical knowledge
- Precision medicine
- Predictive health modelling

Challenge: variety of sources

1. Clinical data

Challenge: variety of sources

1. Clinical data
2. Research data

Challenge: variety of sources

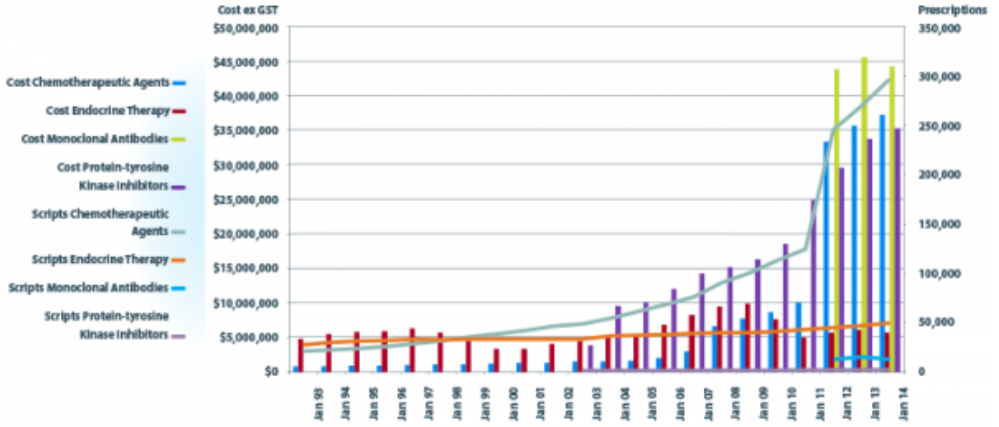
1. Clinical data
2. Research data
3. Patient generated

Challenge: variety of sources

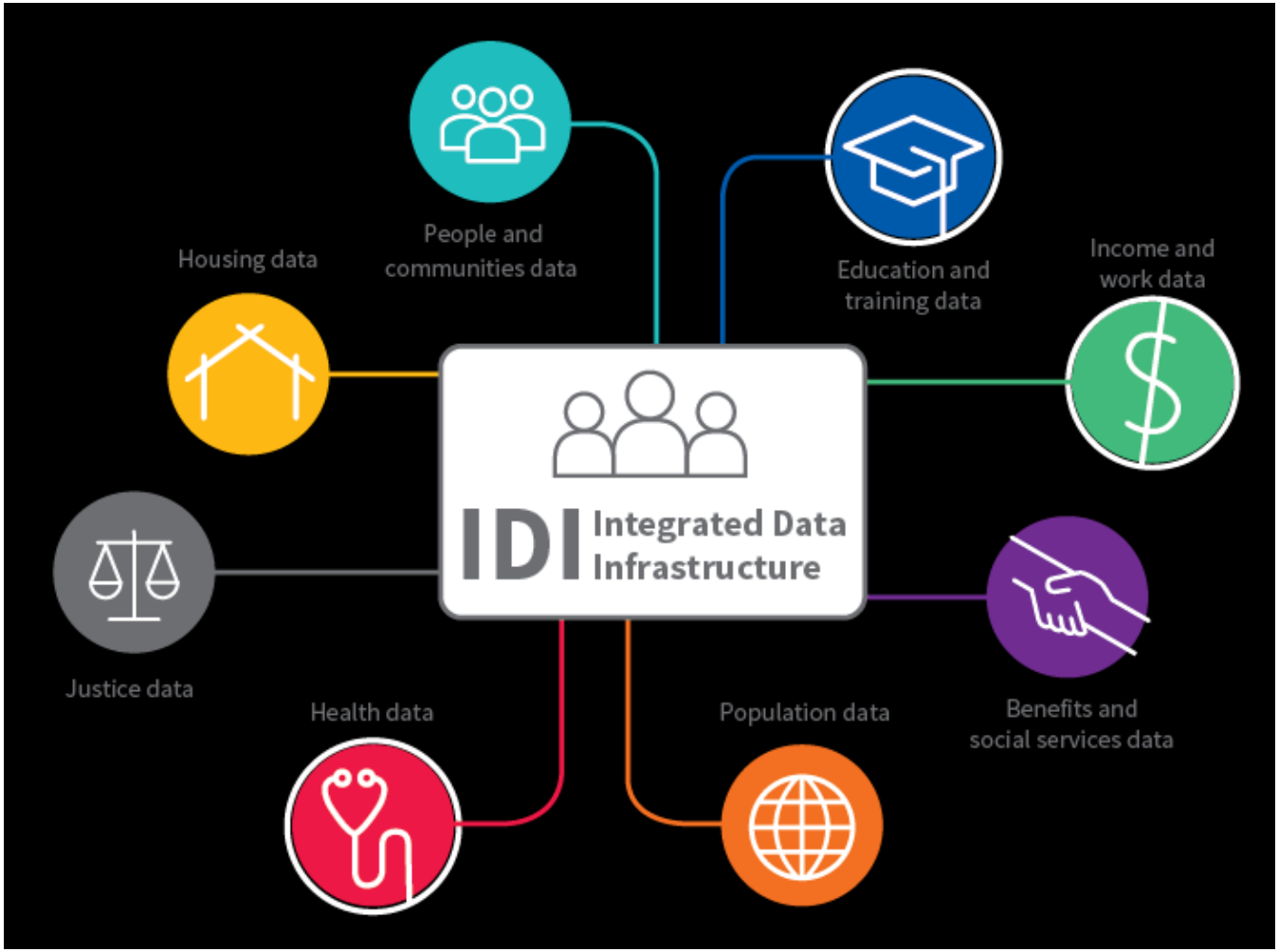
1. Clinical data
2. Research data
3. Patient generated
4. Laboratory

Challenge: variety of sources

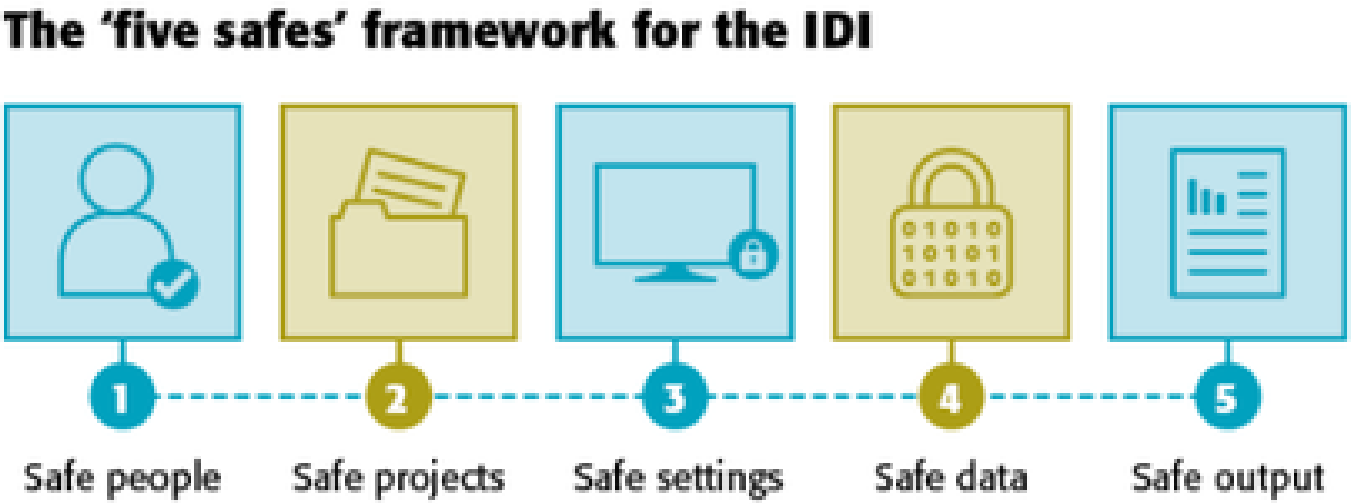
1. Clinical data
2. Research data
3. Patient generated
4. Laboratory
5. Administrative



IDI (NZ)


People and communities data
Housing data
Education and training data
Income and work data
IDI Integrated Data Infrastructure
Justice data
Health data
Population data
Benefits and social services data

IDI governance framework



Treasury 2015: Children at risk

Four key indicators of higher risk – Children aged 0 to 14

Using information collected by government agencies we can identify four indicators that are associated with having poor outcomes later in life. These are:

INDICATOR 1

Having a CYF finding of abuse or neglect

8% of children

INDICATOR 2

Being mostly supported by benefits since birth

15% of children

INDICATOR 3

Having a parent with a prison or community sentence

17% of children

INDICATOR 4

Having a mother with no formal qualifications

10% of children

Data sharing controversies

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Government demands private data from NG

RNZ

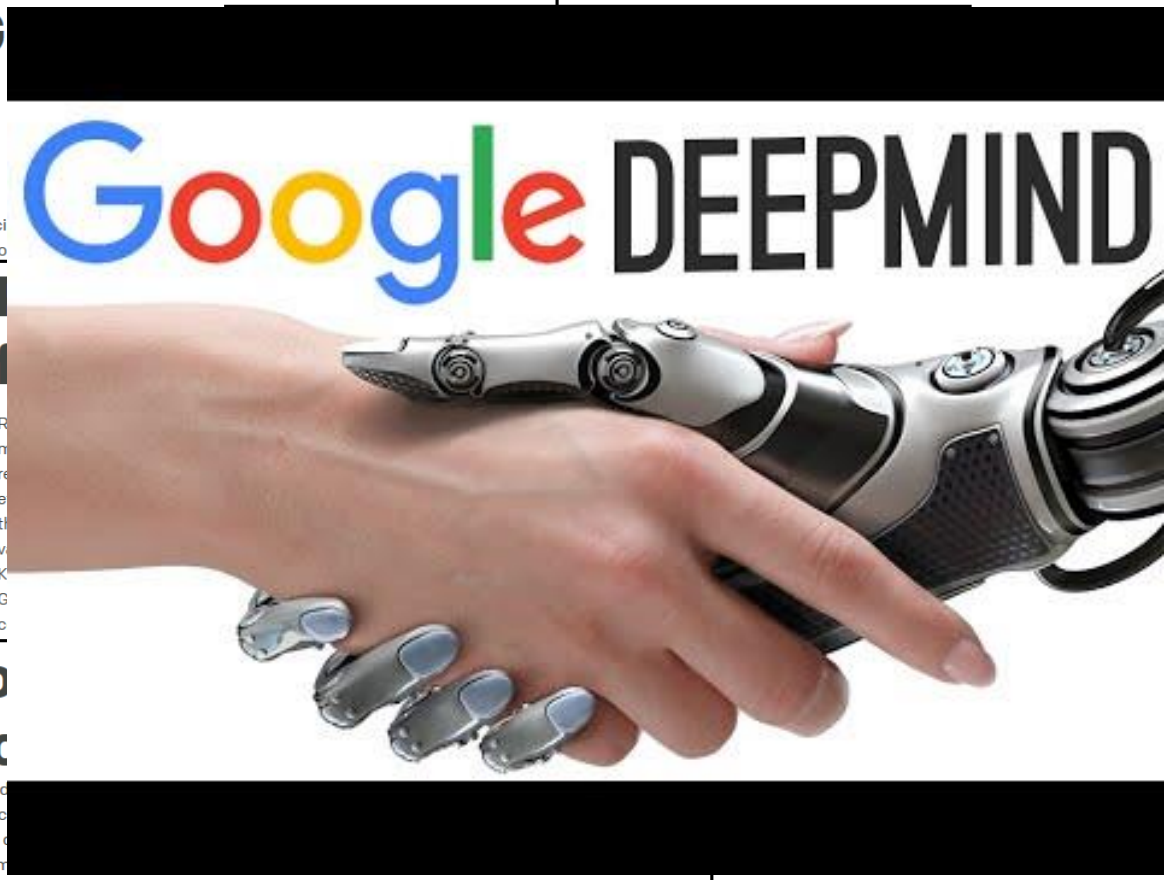
From **Nine To Noon**, 9:09 am on 2 March 2017

Listen 9′ 23″
Add to playlist | Download

Community NGOs say the government is forci
exchange for funding. Kathryn Ryan speaks to

SECTION

Vibra
comp
Published:



Tempo
data fo

The government will d
the organisation whic
Anne Tolley, has now
sexual violence victim
Pilot is from ComVoices, a group which represents the NGO sector.

through the accompanying app, which allows users to control the device remotely

Correlation versus causation

Correlation versus causation

Genetic basis for IQ

SECTIONS The New York Times SUBSCRIBE NOW LOG IN

U.S.

First Gene to Be Li
Intelligence Is Repo

By NICHOLAS WADE MAY 14, 1998

Psychologists using a new g
with high intelligence. The
several more of the many g
personality. If the work is c
first time that a gene contri

Swatermark-text
Swatermark-text

NIH Public Access
Author Manuscript
Psychol Sci. Author manuscript; available in PMC 2012 November 15.

Published in final edited form as:
Psychol Sci. 2012 November 1; 23(11): 1314–1323. doi:10.1177/0956797611435528.

Most Reported Genetic Associations with General Intelligence Are Probably False Positives

Christopher F. Chabris[1,*], **Benjamin M. Hebert**[2], **Daniel J. Benjamin**[3], **Jonathan P. Beauchamp**[2], **David Cesarini**[4,5], **Matthijs J.H.M. van der Loos**[6], **Magnus Johannesson**[7], **Patrik K.E. Magnusson**[8], **Paul Lichtenstein**[8], **Craig S. Atwood**[9,10], **Jeremy Freese**[11], **Taissa S. Hauser**[12], **Robert M. Hauser**[12,13], **Nicholas A. Christakis**[14,15], and **David Laibson**[2]

[1]Department of Psychology, Union College

[2]Department of Economics, Harvard University

[3]Department of Economics, Cornell University

[4]Department of Economics, New York University

[5]IFN-Research Institute for Industrial Economics, Stockholm

[6]Erasmus School of Economics, Rotterdam

[7]Stockholm School of Economics

[8]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm

[9]Department of Medicine, University of Wisconsin-Madison Medical School

[10]Veterans Administration Hospital, Madison, Wisconsin

[11]Department of Sociology, Northwestern University

Making sense of data – mixed methods

- E.g. Facebook (2010/11) IRB approved, full consented survey 789 users, questionnaire about friends IRL versus Facebook relationships

Data ethics challenges

- Individuals
  - (identifiable): privacy/confidentiality, surveillance, harms
- Collectives
  - (non-identifiable): stigma, surveillance, entrenching inequality
- Challenge the general assumption that anonymization is sufficient to prevent harm

Control

Consent

- Specific consent (WMA Declaration of Taipei)
- Broad consent (PHO, Future Unspecific Research)
- Dynamic consent (PEER, Genetic Alliance)
- Opt-out (National Statement, Aus)

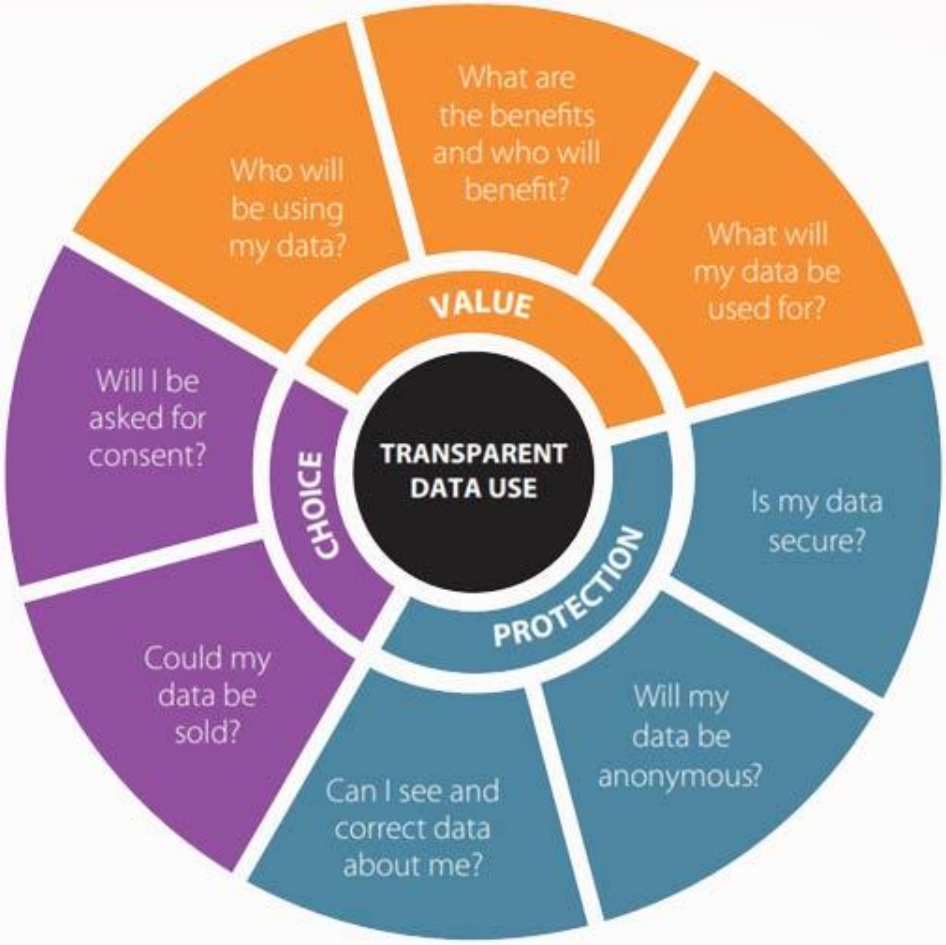
- Problems:
  - Can’t anticipate future data uses
  - Consent from individuals not sufficient
  - Is it necessary?

Governance

- Research ethics committees
- Data access committees
- Data ombudsman (e.g. Dame Fiona Caldicott, the National Data Guardian at the Department of Health)
- Te mana raraunga (cultural licence)

TE MANA RARAUNGA
Māori Data Sovereignty Network

Transparency



Data Futures Partnership

Summary

- Law, ethics, and social licence
- Anonymization is insufficient to prevent harm
- Consider both individual and collective harms