

# U

# O

# W

## Big questions about big data: Social and ethical implications of using real-world data in the public and private sectors

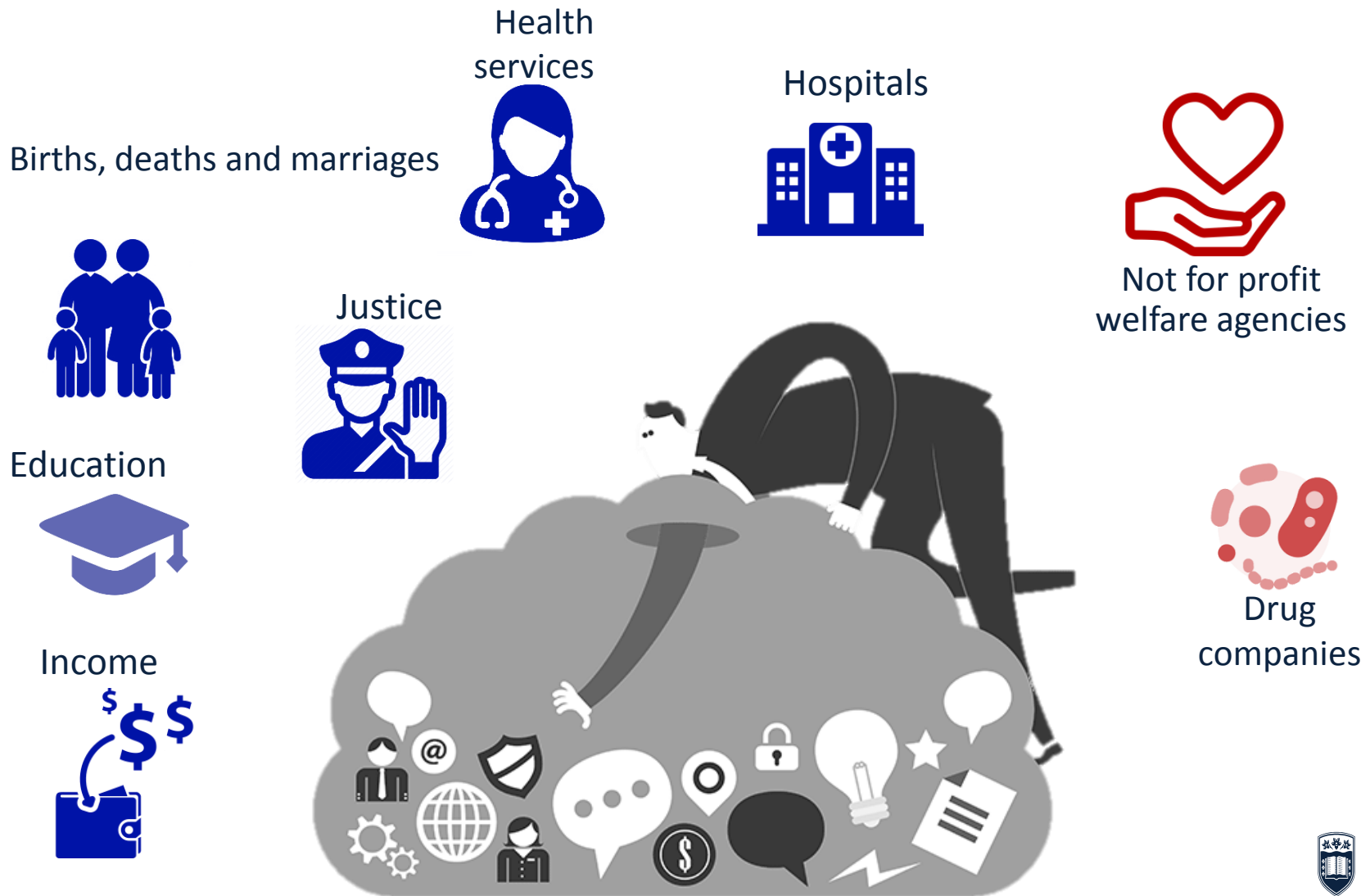
ANNETTE BRAUNACK-MAYER

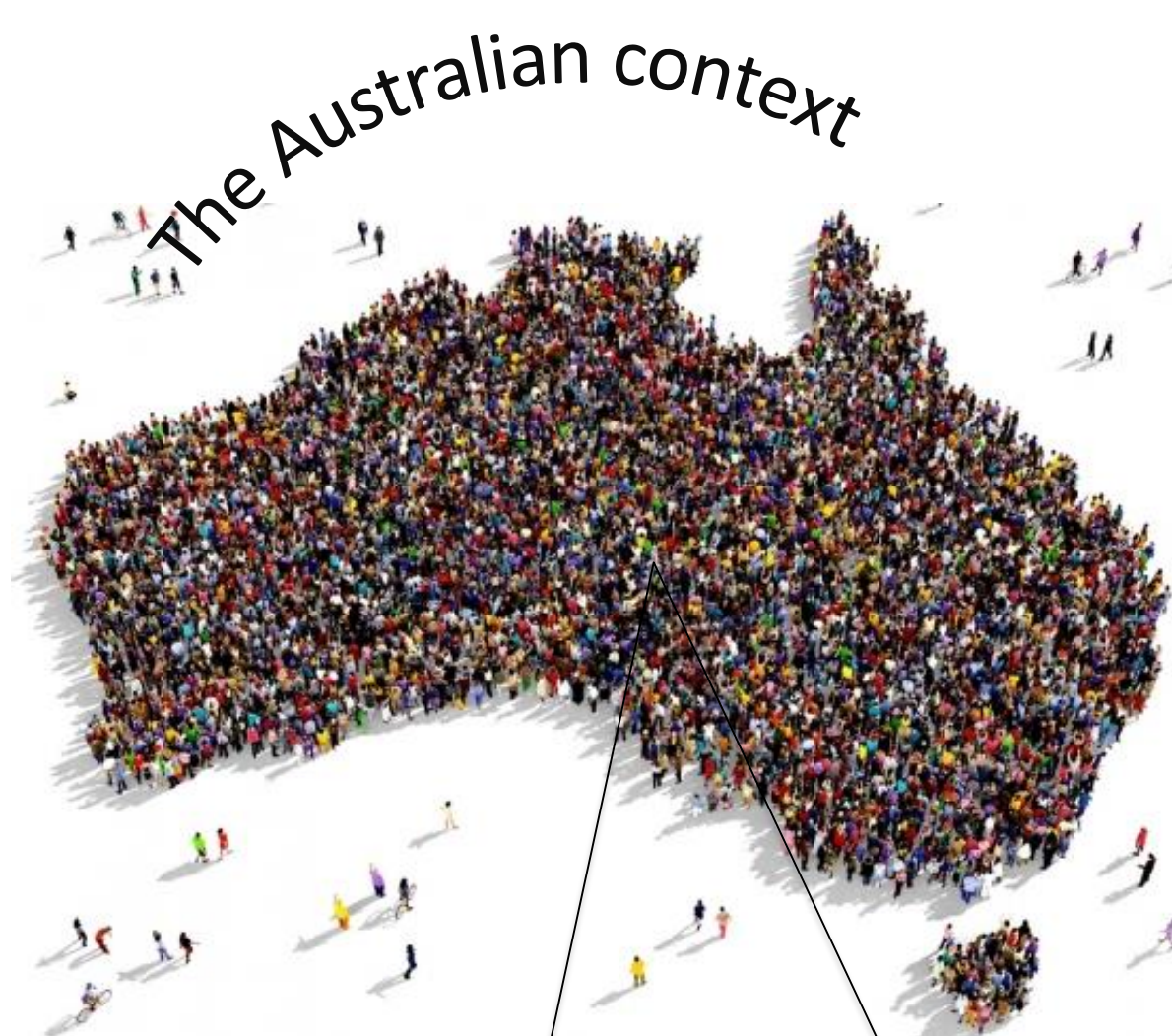
HEAD, SCHOOL OF HEALTH AND SOCIETY



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

# Real-world big data





Community perspectives on big data - including  
in health research



# Big data in universities



# Big data in universities



- Staff and student perspectives
- Ethical issues





# Big data in the Australian context

- Privacy Acts (Commonwealth and State) regulate the collection, use and disclosure of individuals' information
- Apply to both public and private sectors
- Apply to personal information:
  - “Information or an opinion about an identified individual, or an individual who is reasonably identifiable:
    - a) whether the information or opinion is true or not; and
    - b) whether the information or opinion is recorded in a material form or not.”



# Commonwealth Privacy Act

- Act specifies obligations for collectors of data, which are laid out in Privacy Principles, which include:
  - inform individuals of what information is collected and how it will be handled
  - only use personal information for the purpose for which it was collected
  - take reasonable steps to delete or de-identify personal information when no longer needed



# Commonwealth Privacy Act

- De-identified data generally not considered personal data → Privacy Act does not apply
- But, entities are responsible for ensuring that non-personal and de-identified data cannot be linked to individuals



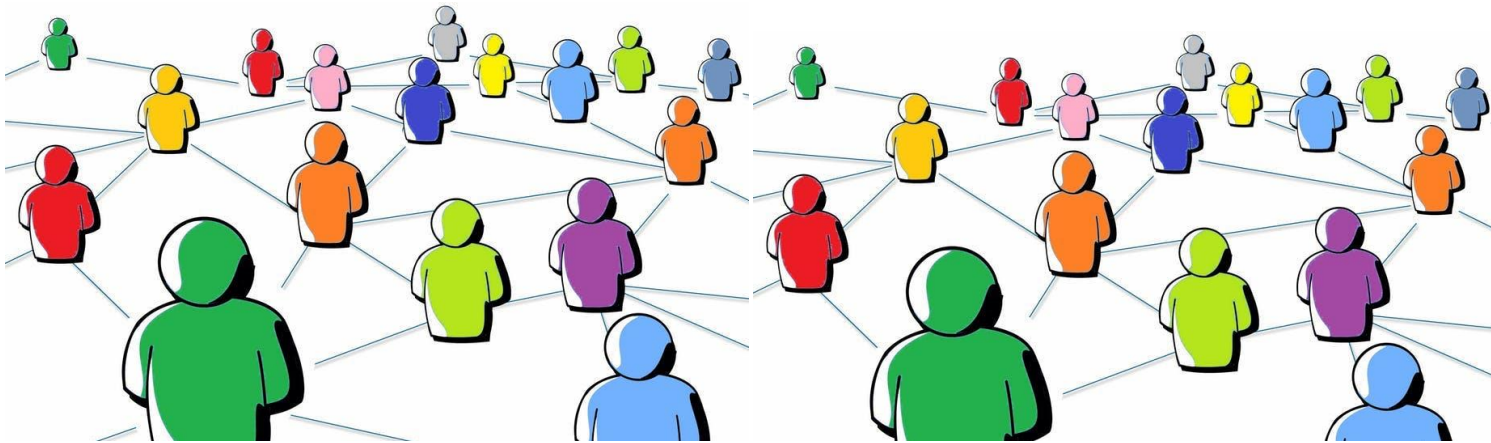
# Human Research Ethics Committees and big data

- HRECs review and approve most research in Australia – including big data research
- Review mechanisms can vary
  - Full review
  - Low risk → streamlined review
  - Negligible risk → exemption
  - National Mutual Acceptance



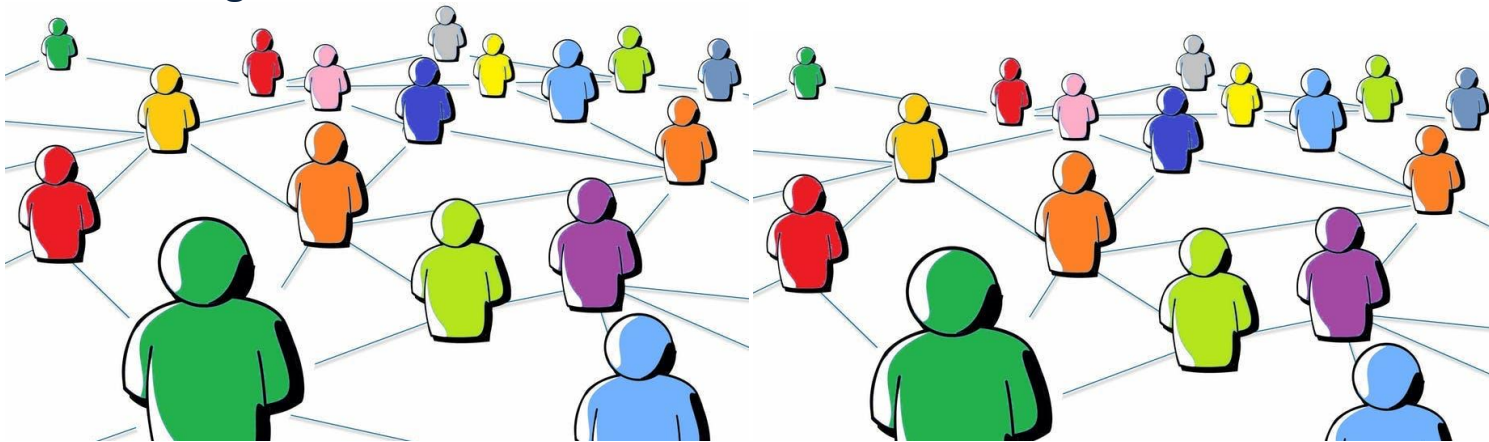
# Human Research Ethics Committees and big data

- Requirements for consent can vary
  - Opt-in
  - Opt-out
  - Waiver
- “Only an HREC may grant waiver of consent for research using personal information in medical research, or personal health information.” (NS 2.3.9)



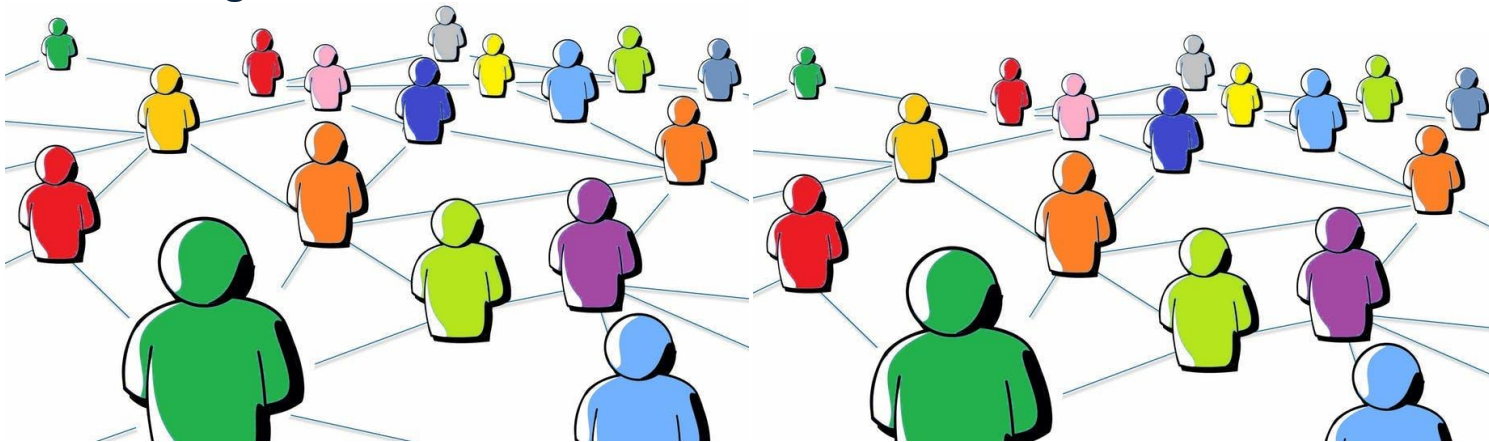
# Human Research Ethics Committees and big data

- Requirements for waiver:
  - Low risk research
  - Benefits > risks when not seeking consent
  - Impracticable to get consent
  - Participants would have said yes
  - Privacy protected
  - Confidentiality of data maintained
  - Outcomes available to participants
  - Participants don't miss out on financial benefits
  - It's legal



# Human Research Ethics Committees and big data

- Requirements for waiver:
  - Low risk research
  - Benefits > risks when not seeking consent
  - Impracticable to get consent
  - Participants would have said yes
  - Privacy protected
  - Confidentiality of data maintained
  - Outcomes available to participants
  - Participants don't miss out on financial benefits
  - It's legal



# Big data in the Australian context - not doing as well as we could



“Australia’s provision of open access to public sector data is below comparable countries with similar governance structures, including the United States, the United Kingdom and New Zealand.”

“Due to a multitude of legal, institutional and technical reasons, Australia stands out among other developed countries as one where health information is poorly used.”

Productivity Commission. (2017) *Data Availability and Use*,  
Report No. 82 Canberra.





# What are the barriers to greater use of big data in Australia?

- Fragile community understanding and trust
- Legislative complexity
- Risk aversion
- Lack of leadership
- Data breaches and re-identification
- Poor quality data

*Source: Productivity Commission. (2017) Data Availability and Use, Report No. 82 Canberra*





# What are the barriers to greater use of big data in Australia?

- Fragile community understanding and trust
- Legislative complexity
- Risk aversion
- Lack of leadership
- Data breaches and re-identification
- Poor quality data

*Source: Productivity Commission. (2017) Data Availability and Use, Report No. 82 Canberra*



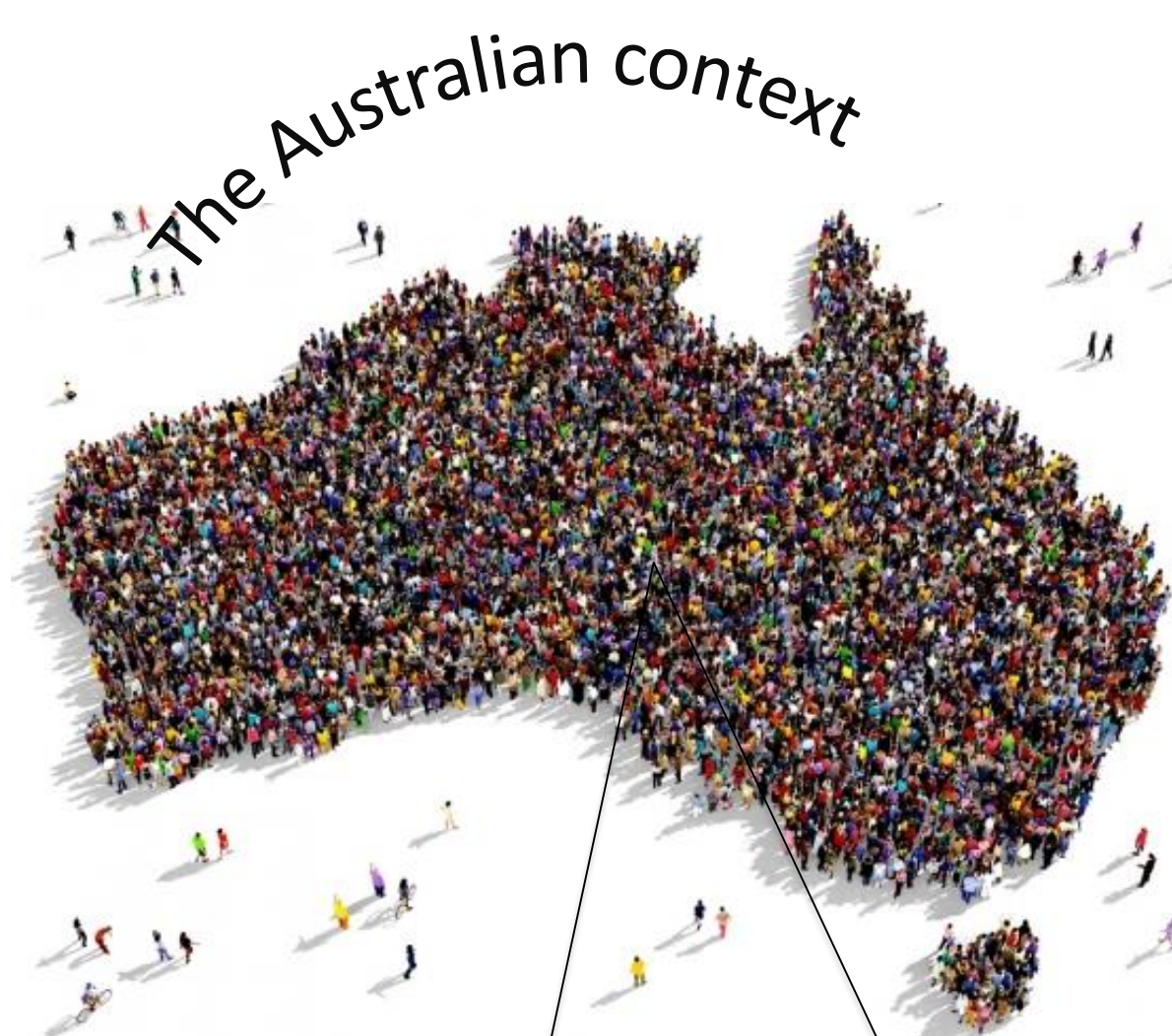


# What are the barriers to greater use of big data in Australia?

- Fragile community understanding and trust
- Legislative complexity
- Risk aversion
- Lack of leadership
- Data breaches and re-identification
- Poor quality data

*Source: Productivity Commission. (2017) Data Availability and Use, Report No. 82 Canberra*





Community perspectives on using our data –  
including health research



# Data breaches

## Census: Australian Bureau of Statistics says website attacked by overseas hackers

Updated 10 Aug 2016, 1:53pm

August 2016

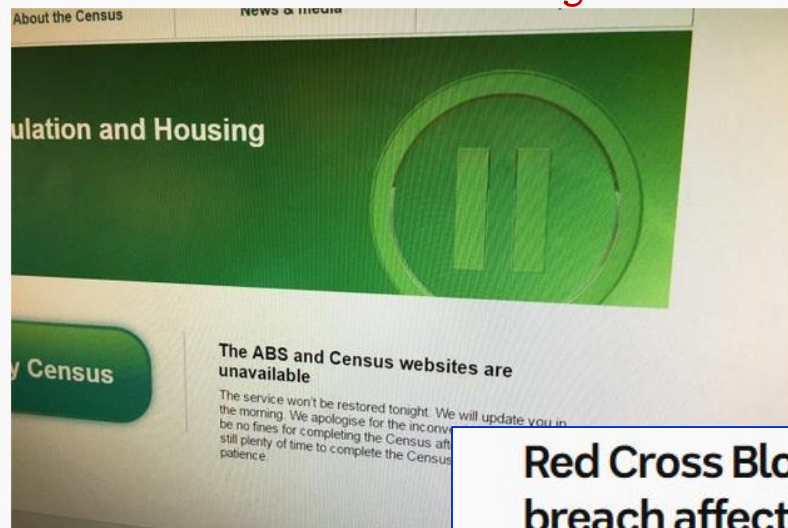


PHOTO: This error message was still showing on the ABS census website

## Not so anonymous: Medicare data can be used to identify individual patients, researchers say

Share on Facebook Share on Twitter

ABC Science By technology reporter Ariel Bogle

Updated 18 December 2017 at 2:20 pm

First posted 18 December 2017 at 1:22 pm

December 2017



## Red Cross Blood Service admits to personal data breach affecting half a million donors

Updated 28 Oct 2016, 5:41pm

October 2016

The personal data of 550,000 blood donors that includes information about "at-risk sexual behaviour" has been leaked from the Red Cross Blood Service in what has been described as Australia's largest security breach.

The organisation said it was told on Wednesday that a file containing donor information was placed on an "insecure computer environment" and "accessed by an unauthorised person".

The file contained the information of blood donors from between 2010 and 2016.

The data came from an online application form and



PHOTO: The file contained the information of blood donors from between 2010 and 2016. (ABC Adelaide: Brett Williamson)



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

# How to find someone in anonymous data...

- 10% sample of Medicare records published in August 2016
- Included gender, birth year, state and health events
- Teague and colleagues:
  - Decrypted IDs of doctors and midwives
  - Identified patients (without decryption) by linking unencrypted parts of record with known information about individuals
- Data pulled down in August 2016



# Searching for myself...

10%  
MBS/PBS  
sample

Gender &  
year of birth

State of  
residence

Number and  
year of birth  
of children



A/Prof Vanessa Teague, School of Computing  
and Information Systems, Melbourne University



# Australians are increasingly sharing information about themselves



7 in 10 Australians use social media



8 in 10 participate in customer loyalty programs



5 in 10 are not fully aware of what data is collected about them and how

## We have a poor understanding of risk



5 in 10 provide false information to obtain online services



4 in 10 read privacy policies



1 in 10 avoid social media due to privacy or security concerns

Source: Directivity et al (2015); ACMA (2012, 2013); OAIC (2013)

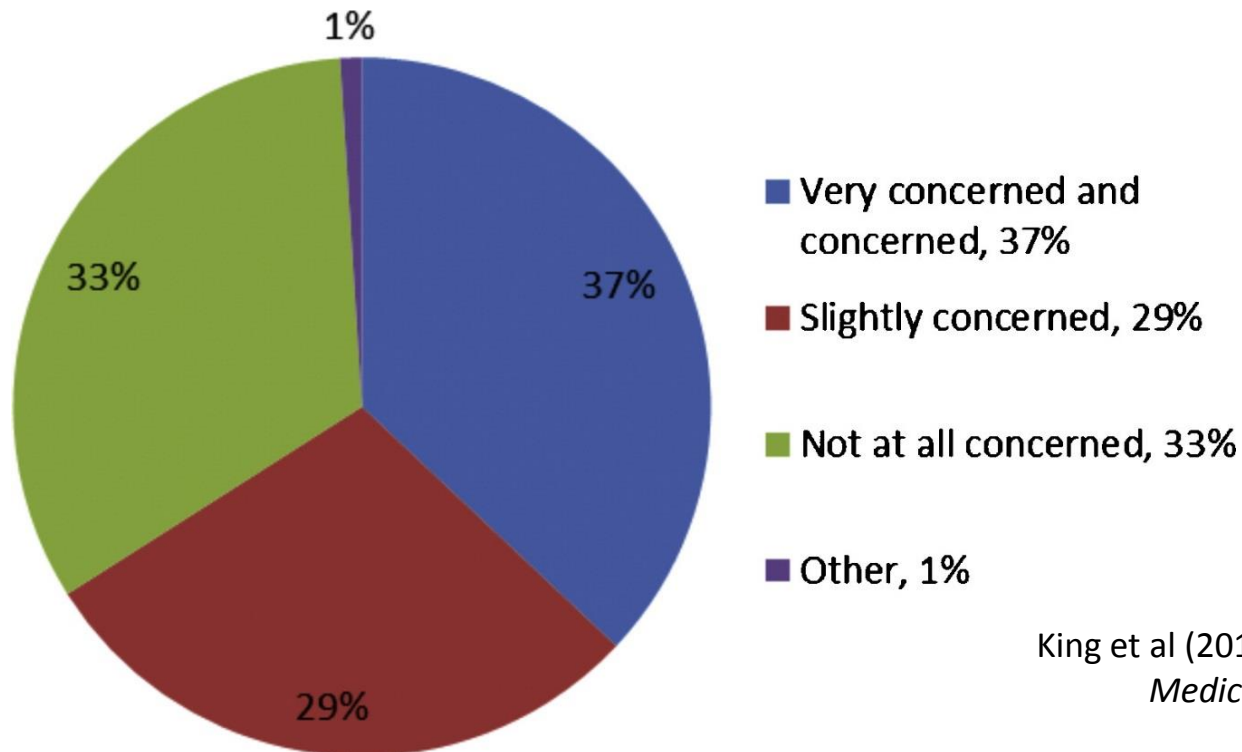


UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

# But, we are worried about privacy and consent...

Health records used for research purposes are generally anonymous, but it may still be possible to link medical information with a patient's name. Knowing this, how concerned would you be about sharing your health information for research?

(n=700)

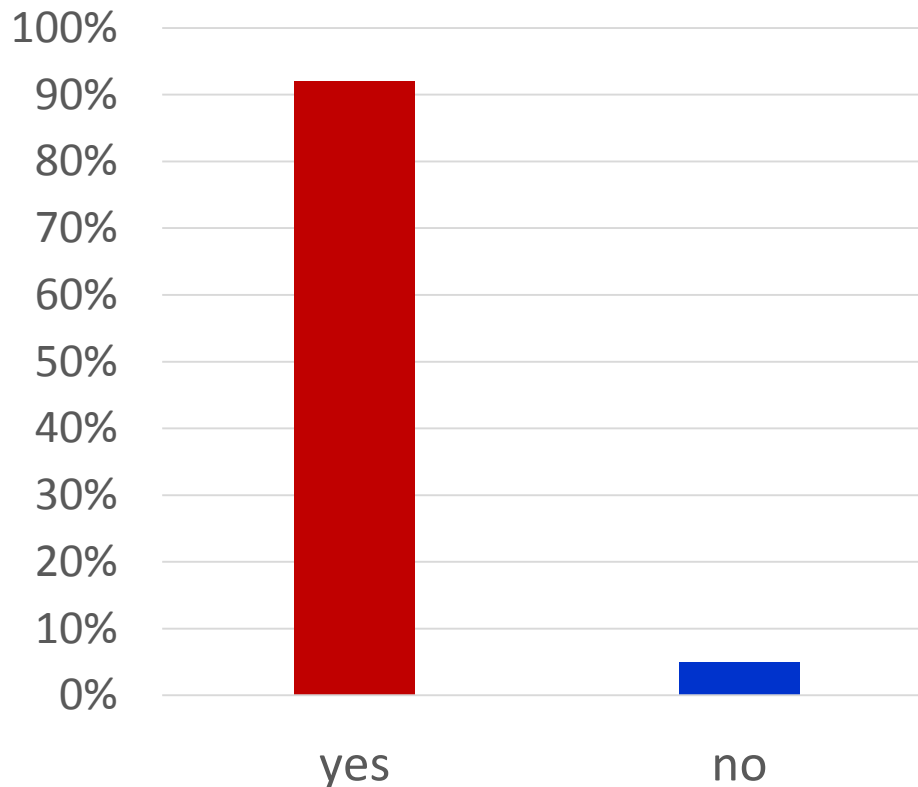


King et al (2012) *International Journal of Medical Informatics* 81: 279–289



# Consent concerns

Do you believe that a person's permission should be sought before their health information is used for any purpose other than medical treatment?



King et al (2012)



# It's confusing...

- 91% of Australians willing to share their health data for research purposes, specifically to:
  - advance health and medical research (79%)
  - support healthcare providers in improving patient care (74%)
  - assist public health officials in tracking diseases & disabilities (68%)
- But, only 51% supported sharing health data with government organisations

Source: Research Australia (2016)



# My Health Record

My Health Record  
is growing!



- Personally Controlled Electronic Health Record program launched in 2012
- Announcement of conversion to opt-out in May 2018
- Opt-out period: July-October 2018
- Extended to January 2019
- Current participation rate 90.1%



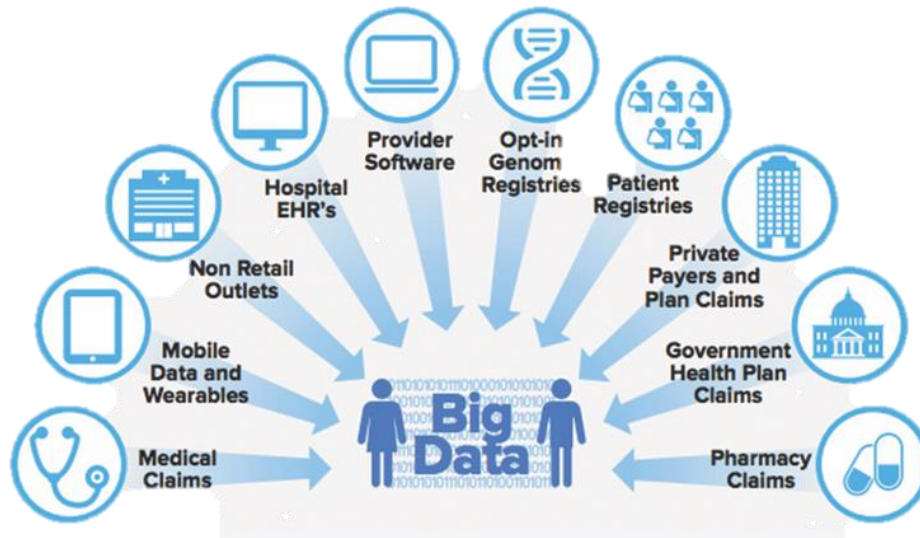
# The privacy paradox in Australia



People claim to be very concerned about their privacy, but do very little to protect their personal data



# What's on the horizon for big health data in Australia?



- New approaches to data availability and use:
  - New data sharing and release framework
  - Consumer Data Right
  - National Data Commissioner
  - Legislation to streamline data sharing and release



# Summary point: in the health sector

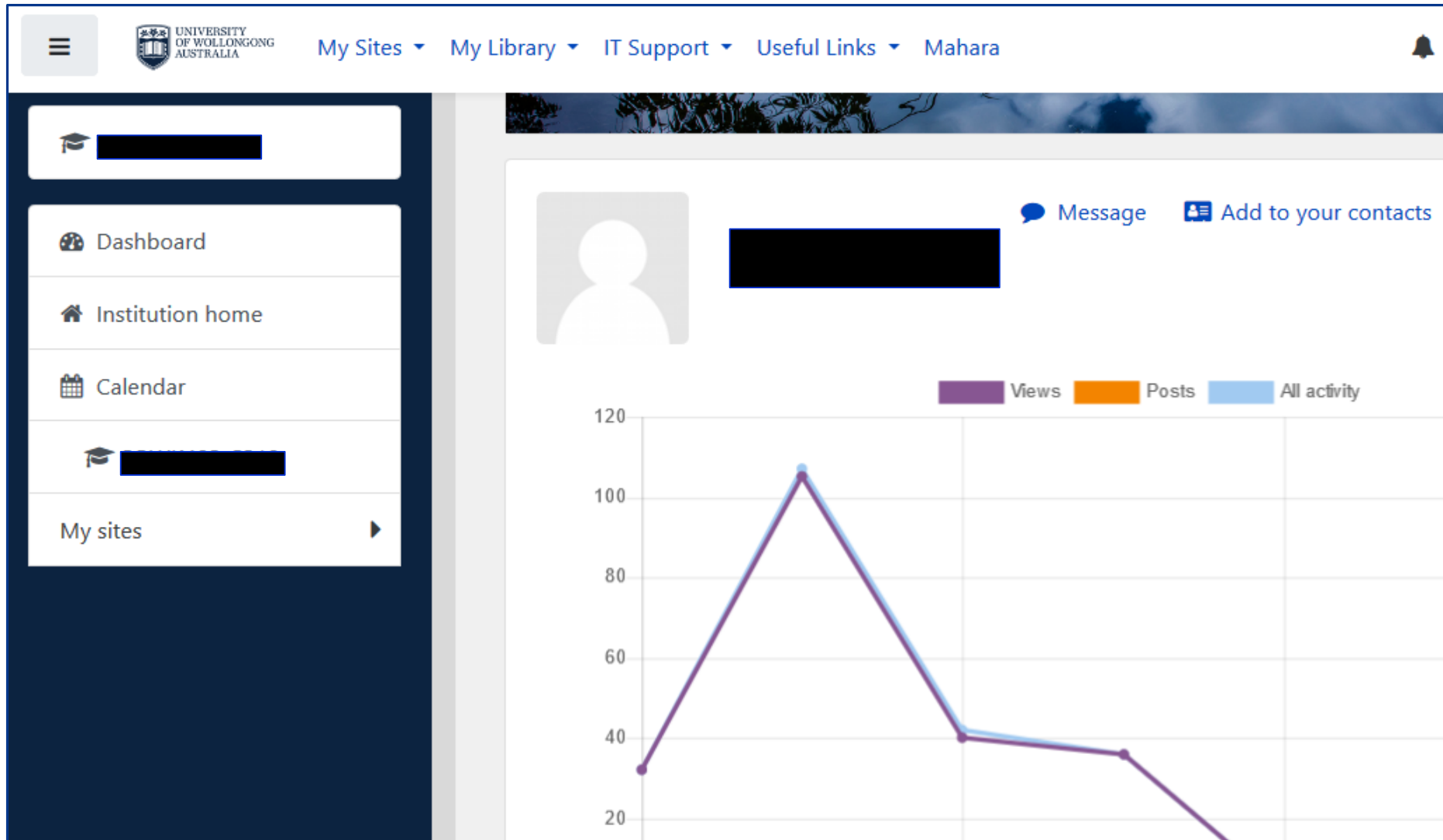
- Community views about use of big data are well documented
- Findings are confusing
- People are becoming less trusting and more anxious



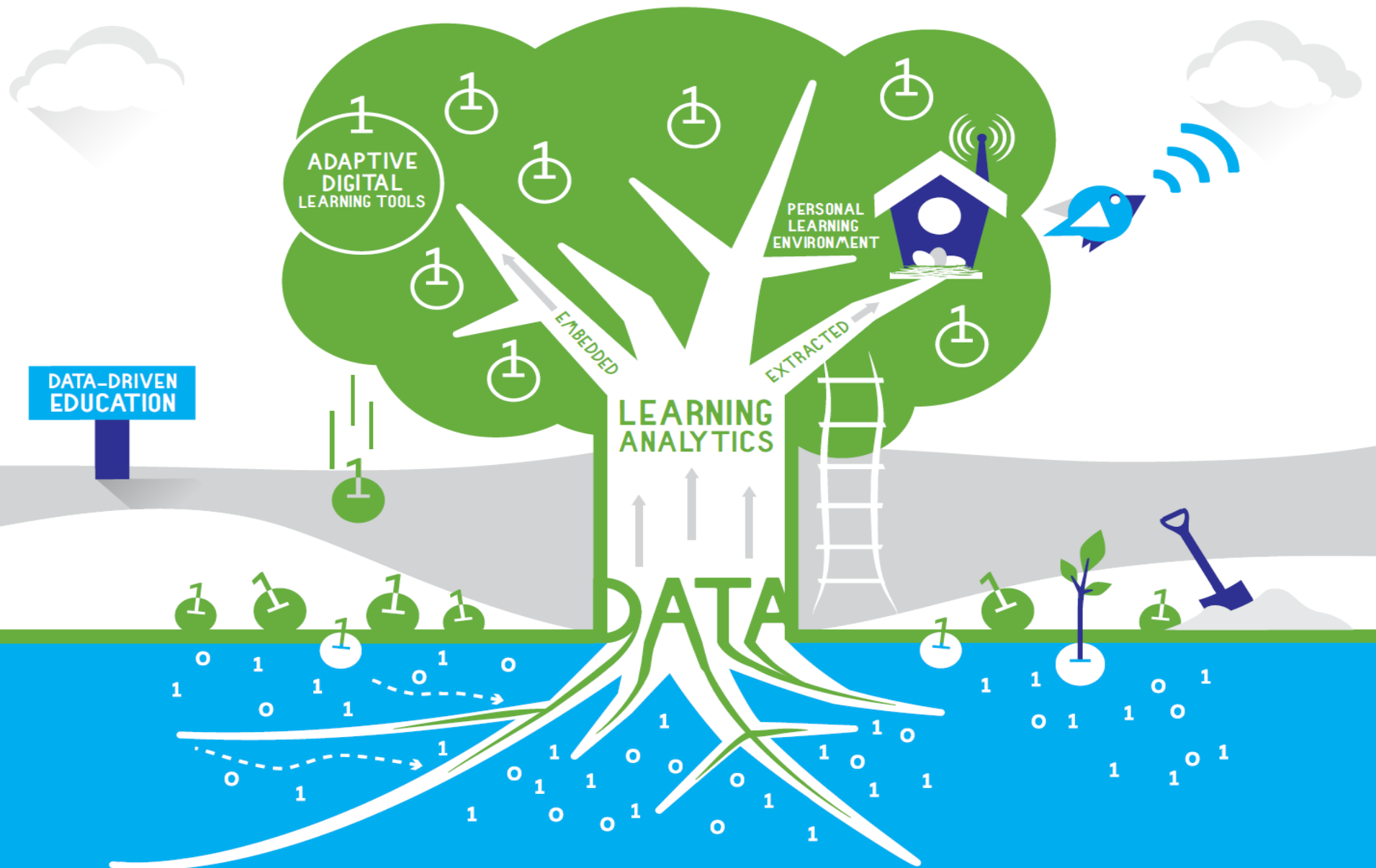
# Big data in universities

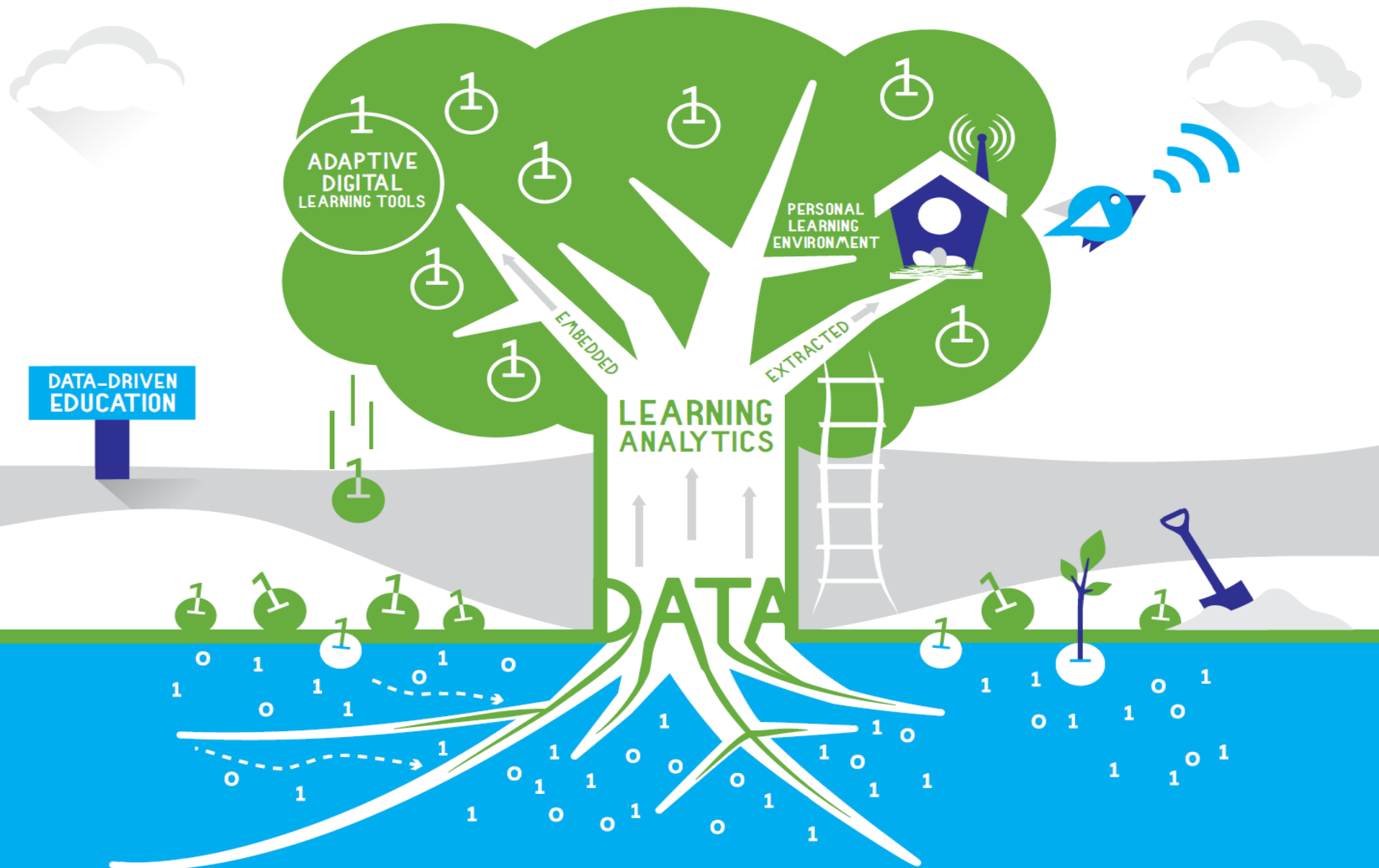


# University learning management systems



# Big data in universities





# Systematic scoping review

# What are the views of students and staff in the university sector on the use of student information in the tertiary sector?

## What ethical issues do these views highlight?



# Databases

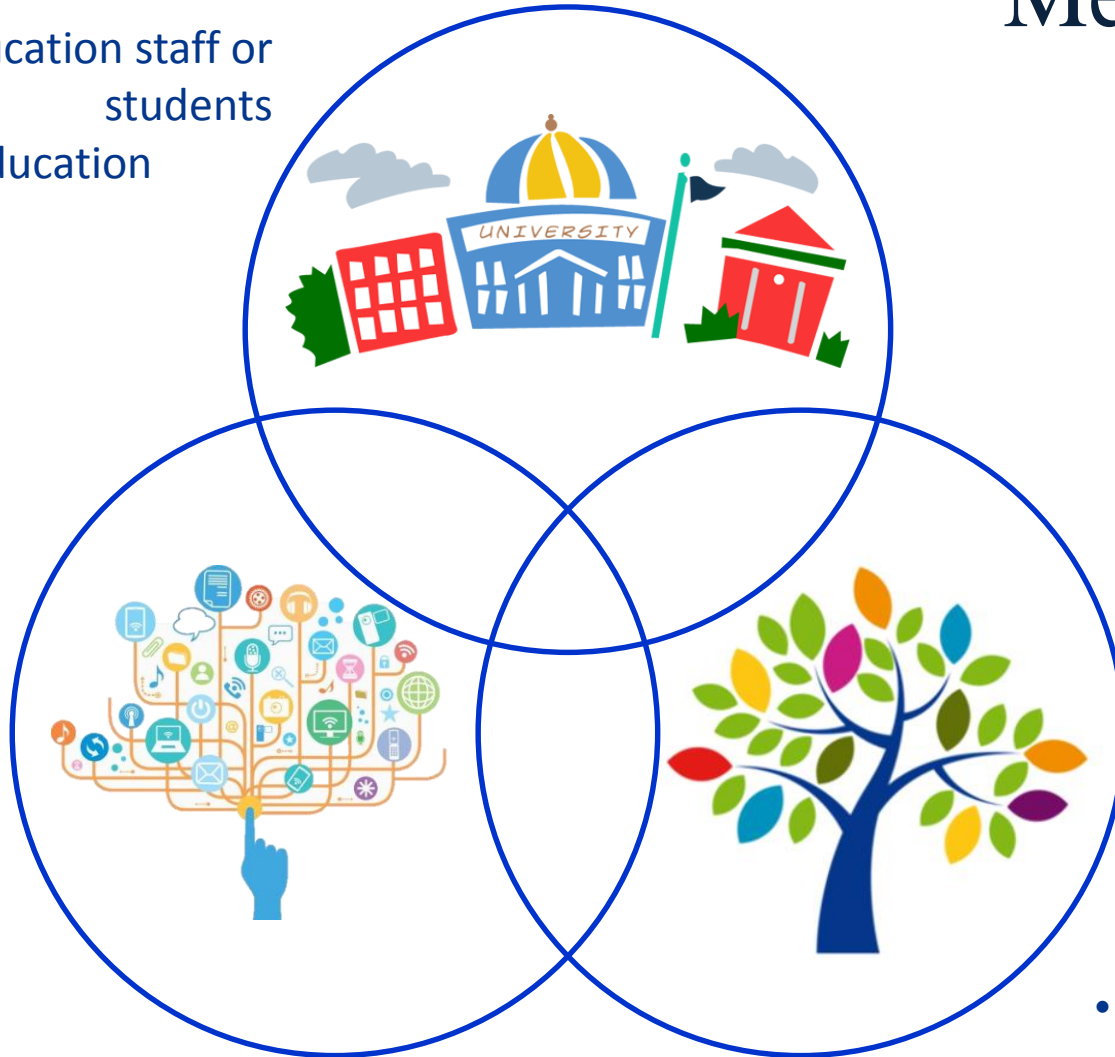
Scopus, ProQuest (ERIC), Education Research Complete, Web of Science, Academic Research Complete, Informit (multiple databases), PsycINFO (Ovid)

Google Scholar (first 1000 hits)



# Methods

- Tertiary education staff or students
- Tertiary education

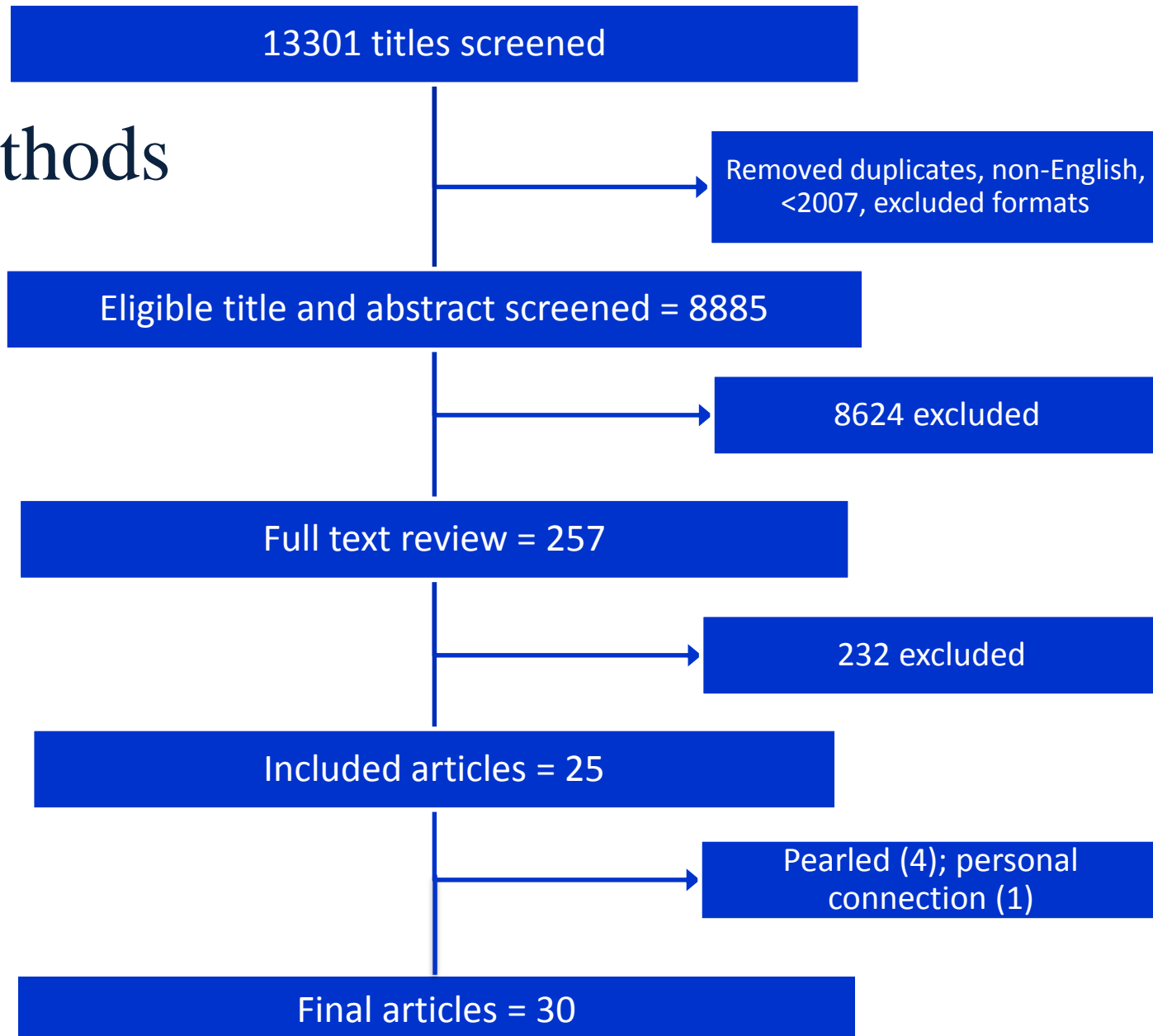


Big data *or*  
Data analytics

- Views *or* perspective
- Ethical issues



# Methods



# Findings

- 30 papers
  - 26 separate studies
  - Student (n=22) and staff (n=11) perspectives
- Main focus – student and staff knowledge, preferences and concerns
- 22 papers acknowledged ethical issues; 8 discussed issues in depth



# Setting

## Student views

- 7 Australia
- 7 UK
- 4 US
- 3 Germany and Australia
- 3 Malaysia
- 1 Germany and France

## Staff views

- 6 Australia and NZ
- 3 US
- 2 International
- 1 UK



# Findings

- Big data in tertiary sector = ‘learning analytics’



- Range of definitions

10 papers: “the measurement, collection, analysis and reporting of big data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs” (Siemens, 2011)



# Findings – awareness

- Both students and staff poorly informed about:
  - Where and what data were stored
  - How data were used
  - Whether guidelines were in place for use of student data
  - Ethical issues
  - Potential future uses using data linkage



# What should big data in universities be used for?

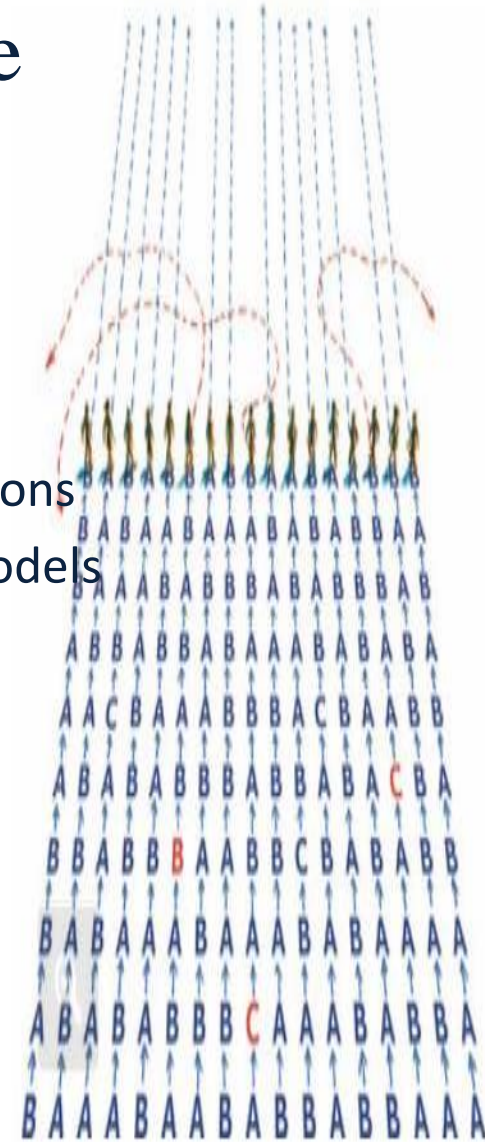
## Student views

For *individual* students:

- Reminders
- Study planning
- Reduce isolation
- Appeal to different learning styles
- Provide helpful feedback
- Identify at risk students
- Monitor performance

## Staff views

- Proactive teaching
- Evidence-based decisions
- Develop predictive models
- Optimise resources



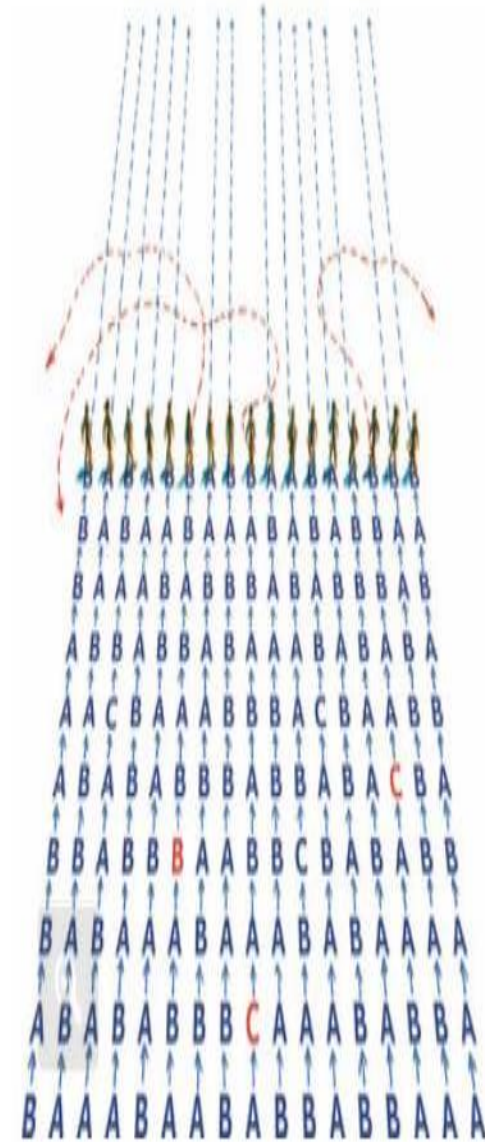
# What are the potential harms?

## Student views

- Anxiety about surveillance
- Privacy intrusions
- Setting unrealistic expectations
- Create a 'helicopter university'

## Staff views

- Inappropriate KPIs
- Increase workload
- Inadequate backup for staff



# Ethical issues

- Privacy and Confidentiality

- Mix of views about how much universities can be trusted to keep information confidential

- Consent

- Recognition that opt-in informed consent for use of student data was difficult
- But, students wanted to be informed and provide consent (perhaps via opt-out mechanisms)

- Transparency

- More information required about what data are collected and how, and privacy protections

- Identity

- Concern that profiling students in certain ways may limit students' future opportunities
- Concerns about inaccurate data and inappropriate algorithms



# What are universities doing about big data?

- Many universities do not appear to have clear policy frameworks
- 8/22 Australian universities yet to consider the ethical implications of big data (West 2016)
- No studies that explored implications of linkage with other datasets



# The University of the Future...

Are certain university degrees associated with higher rates of depression?

- Rationale: Target mental health service provision and develop better intervention strategies for students in need
- Link: Student data such as (course information, enrolment information and learning analytics data) to external data sources (MBS, PBS, hospital emergency attendance, hospital admissions)
- Identifiers (name, DOB, address) used only during linkage stage



# Summary point: in the university sector (in Australia and internationally)

- Low awareness
  - data collection, data analytics, use of predictive analytics
- No awareness of potential for data linkage
- Lack of engagement with students and staff
- Limited understanding of social and ethical issues



# Conclusions



- The storm that is about to hit us
- “A good science fiction story should be able to predict not the automobile but the traffic jam” (Frederik Pohl, 1919-2013)
- ...”and the stranglehold on gas prices”. (Larry Niven, 1938-)



# Acknowledgments

- Jackie Street, Belinda Fabrianesi, Rebecca Bosward and the ACHEEV team
- Lydia Braunack-Mayer
- South Australian Department of Health and Wellbeing
- Bellberry Pty Ltd

