

Data linkage: unlocking the benefits for the VET sector



Presenters



Kristen Osborne
Graduate Research Officer
NCVER



Cain Polidano
Senior Research Fellow
Melbourne Institute of Applied
Economic and Social Research

Presenters



Patrick Korbel
Quantitative Analyst
NCVER



Tenniel Guiver
Data Linkage Manager
Australian Institute of Family Studies

Data linkage in VET

What is it and why do it?

Kristen Osborne
Graduate Research Officer
NCVER



What is data linkage?

- Data linkage involves connecting data relating to an individual entity from different sources, to form a new dataset
- *Deterministic* data linkage uses unique identifiers to definitively match the data
- *Probabilistic* data linkage uses an algorithm to rank the probability of records belonging to the same entity, then matches those that have a probability of matching above a predetermined threshold

Why link data?

- Data linkage offers the opportunity for enhanced analyses and has the potential to provide valuable multifaceted insights for policy and further research
- Data linkage can be used to capture a more complete picture of an individual's lifelong journey

Privacy in data linkage

- A particularly important issue for data linkage is the protection of people's privacy
- All data linkage projects must consider and ensure privacy compliance.

Privacy in data linkage

Options are available for mitigating the privacy risks:

- Conducting a privacy impact assessment
- Ensuring datasets are appropriately de-identified
- Creating a one-off rather than an enduring linkage
- Establishing limited access to the linked dataset

Other issues in data linkage

- Dealing with various data custodians
- Obtaining consent where appropriate
- Ensuring compliance with all relevant state/territory and local government legislation and policies
- Resourcing costs associated with data linkage
- Security of and access to linked data

1 REVIEW

- current knowledge of the relevant policy purposes.
- the present evidence base that helps explain any of the economic/social/environmental issues that impact the area of intended inquiry.



a

Create an inventory of all relevant quantitative data sources used and pertinent qualitative inquiry, research and reports.

2 STATEMENT & QUESTIONS



Determine a statement of the central research objective, supplemented by a minimum set of research questions.

This may take the form of a testable 'model/hypothesis'.

3 EVALUATE DATA

Evaluate available data or established datasets on the basis of the following factors:



a

Careful consideration of:

- all privacy and/or ethical concerns and risks
- data access, use and permissions needed
- limitations, including conducting a privacy risk assessment.

b

Relevance to research objective and questions, intimacy to the proposed model/hypothesis being tested.

c

Data quality and access, including:

- evidence of longitudinal robustness and reliability
- whether data is publically available
- any resource effort in data assembly.

d

Efficacy, ease and reliability of any chosen means of 'linkage' between any datasets.

e

Whether any data linkage approach is reliable and replicable.

4 PROOF OF CONCEPT

Conduct 'proof of concept' scenario tests to determine the minimum necessary datasets and their optimal linkage.

Assess which approach best addresses the research objective and questions, and has highest likelihood of generating new knowledge and insights.



a Decide on optimal approach based on benefit/cost/time.

b Think through possible confounding 'cause/effect' interpretations or issues that may compromise drawing valid conclusions from the work.

c If appropriate run a smaller 'pilot' study.

5 SET UP STUDY

Conduct in compliance with approved permissions and data security requirements and limited to the purposes that have been consented to by individuals. If possible, make the research results available to individuals within the linked dataset.



6 SECURE

Secure created datasets, making them available within appropriate authorisations.



Data linkage for the VET sector

- In the VET sector specifically, linking data from the National VET Provider Collection (which includes information on VET students, program enrolments, subject enrolments, program completions and source of funding) with:
- Surveys such as the ABS Census, the National Health Survey or the ABS General Social Survey
- Administrative data such as welfare information or tax records
- Proprietary data, including collected by employment organisations such as SEEK
- This can provide more in-depth knowledge of pathways through, outcomes from, and the impact of, VET on the different aspects of an individual's life.

Data linkage or data integration?

A research case study

Cain Polidano

Melbourne Institute of Applied Economic and Social Research

University of Melbourne

More impactful research

- Better understand policy outcomes, e.g.:
 - AVETMISS linked to ATO would help understand long-run outcomes from changes to VET
- Understand how policies across agencies interact, e.g.:
 - AVETMISS linked to employment service provider data can shed light on how the outcomes of VET for jobseekers is affected by employment services
- Understanding policy interaction can improve policy co-ordination and effectiveness
 - Vital for addressing deep-seeded disadvantage

A research example, warts and all

- Evaluation of complementary indigenous employment services provided by **Department A** that are ‘over-and-above’ standard services provided by **Department B**
 - Program 1: matches jobs seekers with partner employers and provides ongoing mentoring
 - Program 2: vocational training in preparation for guaranteed job with partner employers
- Do the complementary services (**A**) improve the chances of attaining sustainable employment compared to standard services only (**B**)?

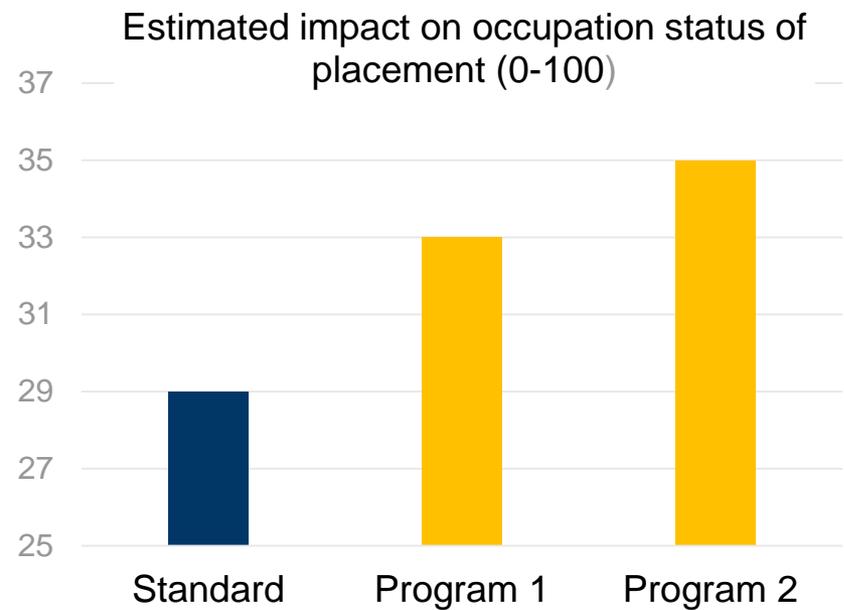
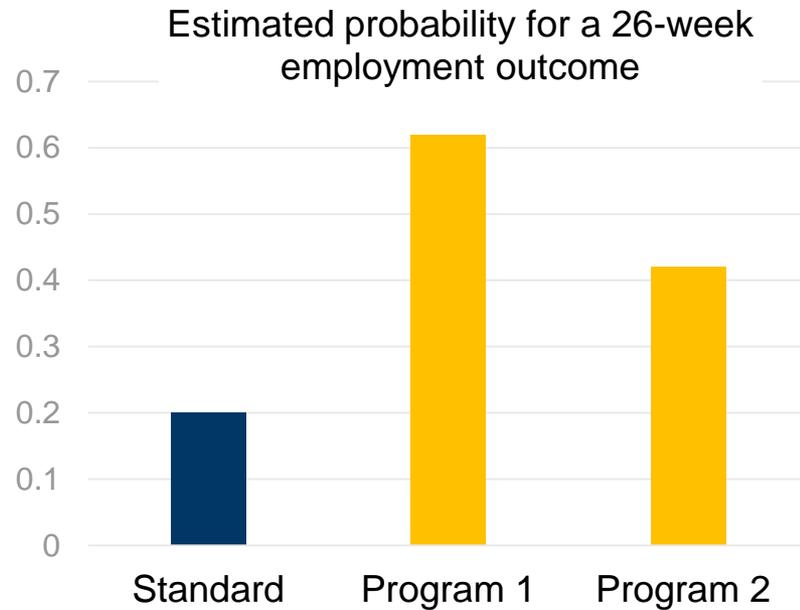
Empirical approach

- Employment outcomes from those who received complementary services (A), relative to a ‘like’ indigenous control group who only received standard services (B)
- Base sample is welfare receipt data and standard employment service records from (B). We link complementary service records from (A)
 - Assume that those without linked complementary service records did not access service

Difficulties with data linkage

- Lack of documentation
 - difficulty interpreting variables
- Different institutional settings hampered data use
 - inconsistent variables, variable names, category coding and treatment of 'missing' between (A) and (B)
- Employment service provider data contained no information on services provided (just placement)
 - Makes interpreting results difficult

Results



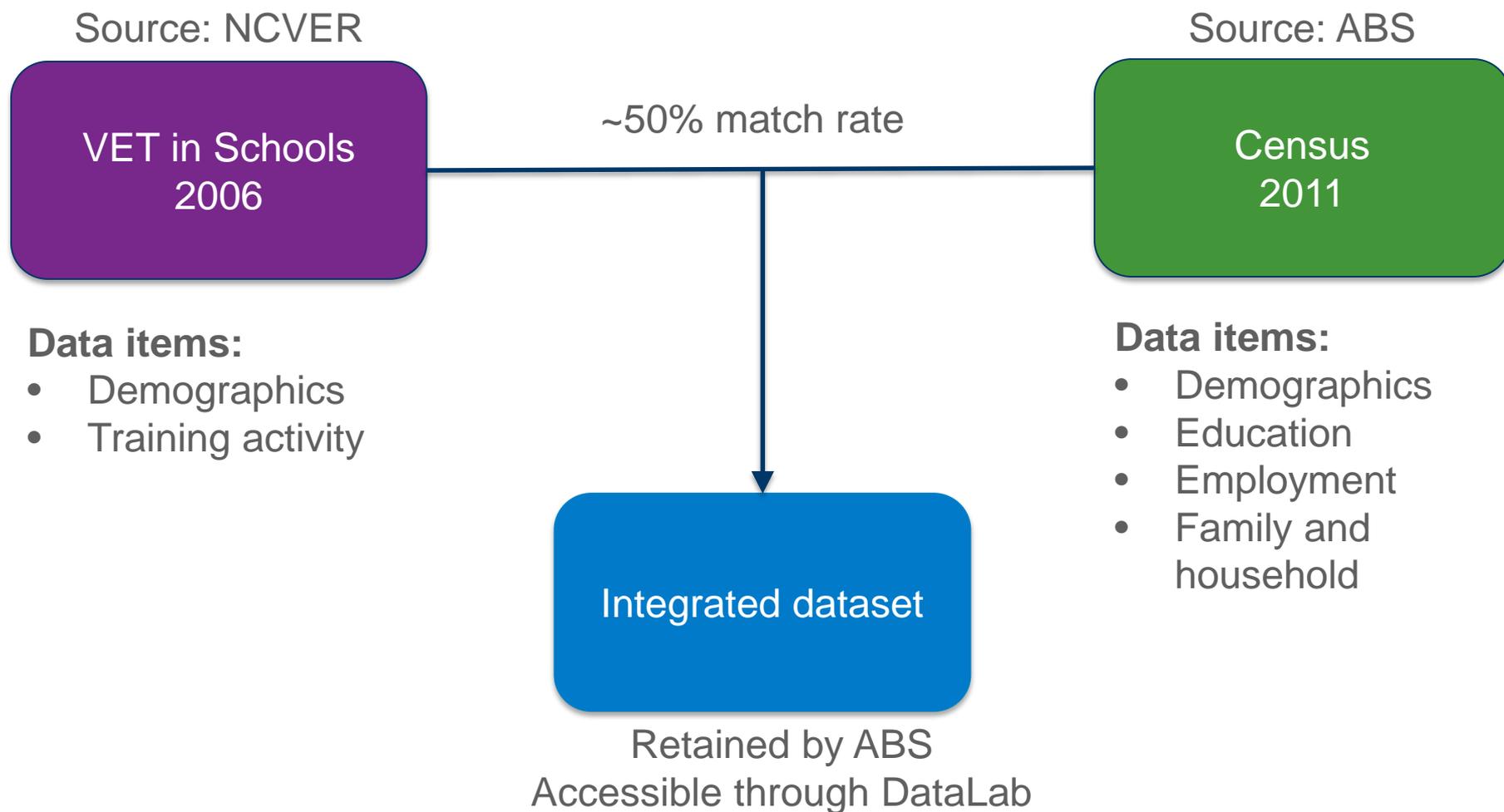
Conclusions

- Data linkage opens-up opportunities to better understand policy interactions
- However, it is more than joining data, it's also about 'data integration'. This requires:
 - Co-ordination across departments
 - Collaboration with data users to define linkage purposes

Linking VET data with the Census

Dr Patrick Korbel
Quantitative Analyst, NCVER

Model of the linkage



Why link the data?

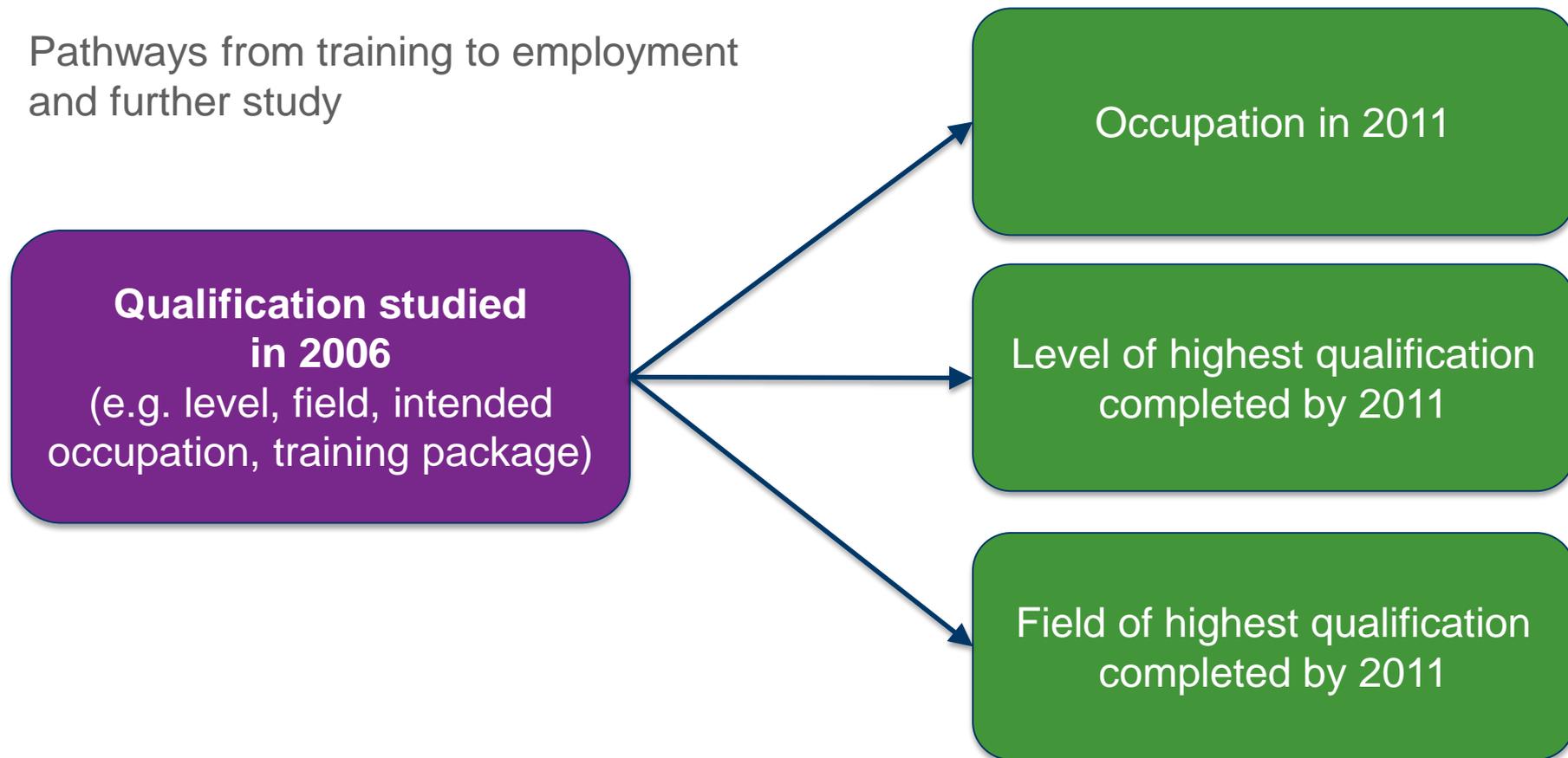
- Capturing student experiences after VET in Schools
- Making effective use of existing data
- Complements the use of surveys such as the Longitudinal Survey of Australian Youth (LSAY)

Characteristics of the linkage

- Low match rate due to data limitations
- Provides a cross-section snapshot of two points in time
- Some activity between 2006 and 2011 is either unknown or inferred

Insights from the data

Pathways from training to employment
and further study



Reflections on the study

- Three pathways identified:
 - Specific vocational preparation
 - General vocational preparation
 - General education and interest
- Addresses data gap and directs further research
- Future studies should involve a ‘control’ group
- This involved two population datasets, but it would be very different linking a survey

Data Linkage at AIFS

... helping paint the bigger picture



Australian Government

Australian Institute of Family Studies

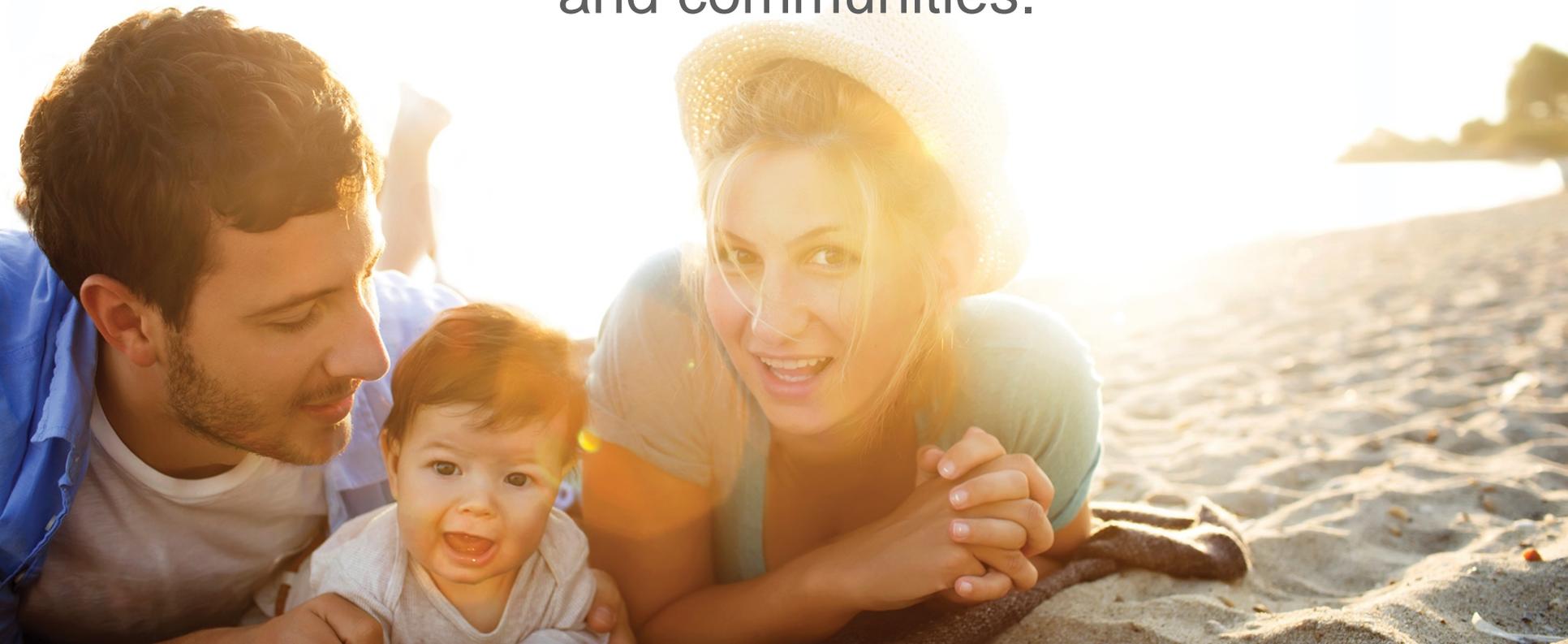
Tenniel Guiver

Data Linkage Manager

AIFS

AIFS Purpose

We create and communicate knowledge to accelerate positive outcomes for families and communities.



Answering research and policy questions



How can we assess educational factors over lifetimes?

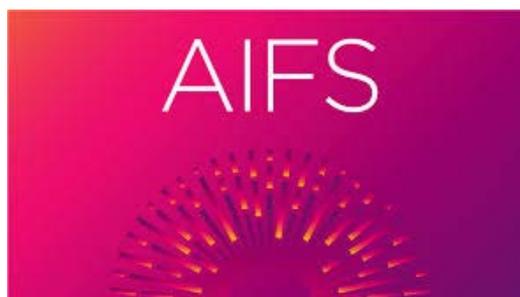
What pathways, policies, and programs lead to the best outcomes?

How can we observe transitions through the education system?

AIFS Data Linkage Capabilities



Integrating
data



Data
scoping



Analysis and
research using
linked data





- Medical services (MBS)
- Pharmaceutical Services (PBS)
- Immunisations (ACIR)

My School[®]

- School performance and profile



- Area based socio-economic indicators



NAPLAN

NATIONAL ASSESSMENT PROGRAM
Literacy and Numeracy

- School literacy and numeracy results



Discovering what works for families



centrelink

- Income support payments
- Family tax benefits payments

Australian Institute of Family Studies

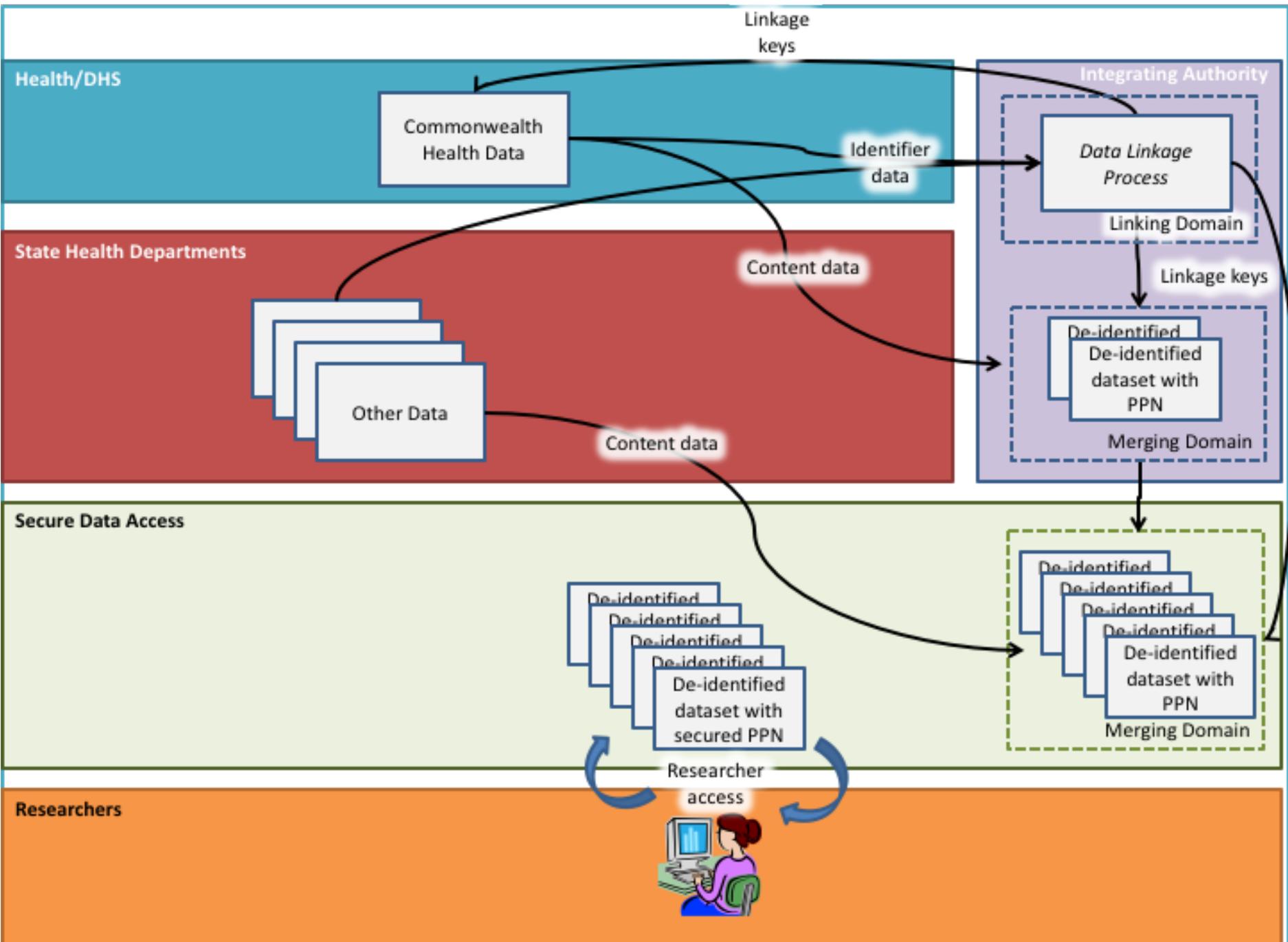
Linkage Methodologies

- Probabilistic
 - Full name, date of birth and sex
 - Address/postcode
 - Date of last contact
- Deterministic - e.g. Medicare Person ID Number, Tax File Number
- ‘Key-based linkage’
 - Community services statistical linkage key (SLK851)
 - Additional information, e.g. Postcode
- Event-based
 - Dates of service provision
 - Date of birth and sex
- Hash keys
- Machine learning

Linkage Challenges

Critical elements - security, capacity, speed, accuracy and cost effectiveness:

- Protecting **Confidentiality & Security** – guarantee to protect confidentiality where data collected & shared.
- Balancing the needs of users against the **burden** on providers
- **Maximise the value** of available administrative data and systems.
- **Obligations and undertakings** associated with each original source.



Protecting confidentiality with the “5 safes”

Five guiding principles:

- **Safe people** – trained accredited researchers, confidentiality undertakings
- **Safe projects** – research projects for ‘public good’, ethics approvals
- **Safe settings** – secure data access mechanisms
- **Safe outputs** – controlling statistical disclosure
- **Safe data** – de-identified and confidentialised data



OVERVIEW



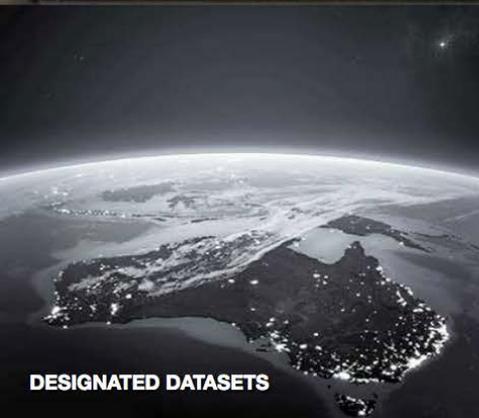
THE CONSUMER DATA RIGHT



GOVERNING THE NATIONAL DATA SYSTEM



ACCREDITED DATA AUTHORITIES



DESIGNATED DATASETS



IMPROVING AUSTRALIA'S CURRENT DATA SYSTEM

NEW LAWS TO IMPROVE ACCESS TO DATA AND USE OF DATA
A Data Sharing and Release Legislative Package

[CONTINUE READING >](#)



REFERENCES

Immediate start

Improving data accessibility and quality

- Create **registries** of public and publicly funded data holdings
- **Publish guidance** for custodians on data sharing and release
- Government agencies to commence implementation of **data standards**
- Commence **releasing all** non-sensitive public sector data
- Designate **agency responsible** for introducing data policy reform
- Appoint **National Data Custodian (NDC)** administratively

Improving data usage

- Accredit more **data linkage units**
- **Abolish requirement to destroy** linked data and linkage keys

Data Sharing and Release Act

- Commence exposure draft development

Consumer rights

- Begin industry negotiations on definition of consumer data
- Monitor progress toward target for Comprehensive Credit Reporting (CCR)

Over the next 12 months

- **Draft Data Sharing and Release Act**
 - draft will include an initial list of National Interest Datasets (NIDs)

- **NDC to be tasked with:**
 - developing a process for accrediting release authorities (ARA) and trusted users
 - developing a process for stakeholders to commence nominations for NIDs
 - consultation on draft Act

- Appoint **advisory board** and ethics adviser to the NDC

- Streamline human research **ethics committees processes**

- Priority development of industry **data specification agreements**

- If required, introduce legislation to **mandate CCR**

- **Report on progress** with implementation of new framework

Reform goals

- Publicly funded entities to publish data registries by mid **2018**
- Data Sharing and Release Act to be passed by end **2018**
- From **2019:**
 - NDC to maintain new data framework
 - ARAs to manage National Interest Datasets
 - Data custodians to work with ARAs to improve data use
 - Comprehensive Right enforced by ACCC
- By **2020**, data standards to be operational across the public sector

Upcoming Events



PROGRAM OUT NOW!

REGISTER HERE
www.ncver.edu.au/nofrills2018

NO FRILLS
27th National VET Research Conference
First international partnership

Skills for a global future
Working and learning together
15-17 August 2018
Sydney

New Zealand
VOCATIONAL EDUCATION AND TRAINING
Research Forum

Upcoming Events



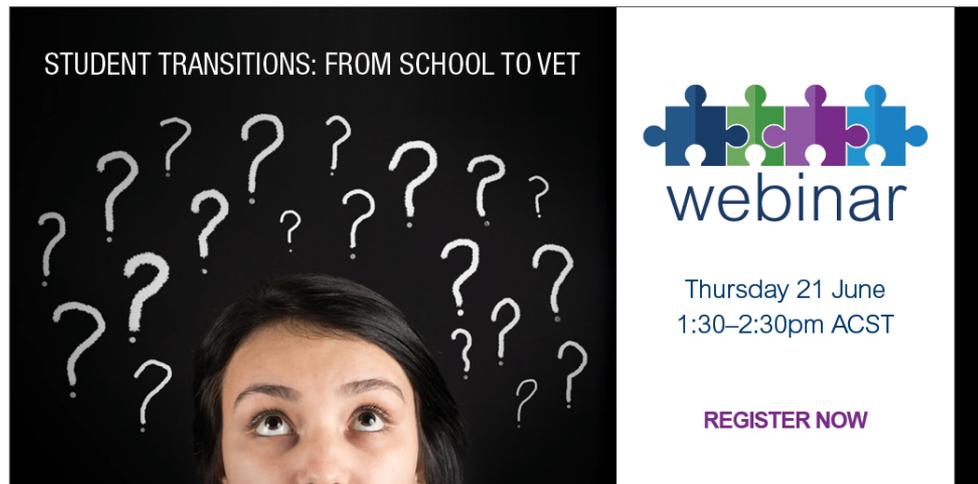
SKILLING FOR DIGITAL DISRUPTION
AND THE FUTURE OF WORK



webinar

Tuesday 19 June
1:30–2:30pm ACST

[REGISTER NOW](#)



STUDENT TRANSITIONS: FROM SCHOOL TO VET



webinar

Thursday 21 June
1:30–2:30pm ACST

[REGISTER NOW](#)

Stay connected

Subscribe
at www.ncver.edu.au

Follow



twitter.com/ncver or [@ncver](https://twitter.com/ncver)



[linkedin.com/company/ncver](https://www.linkedin.com/company/ncver)