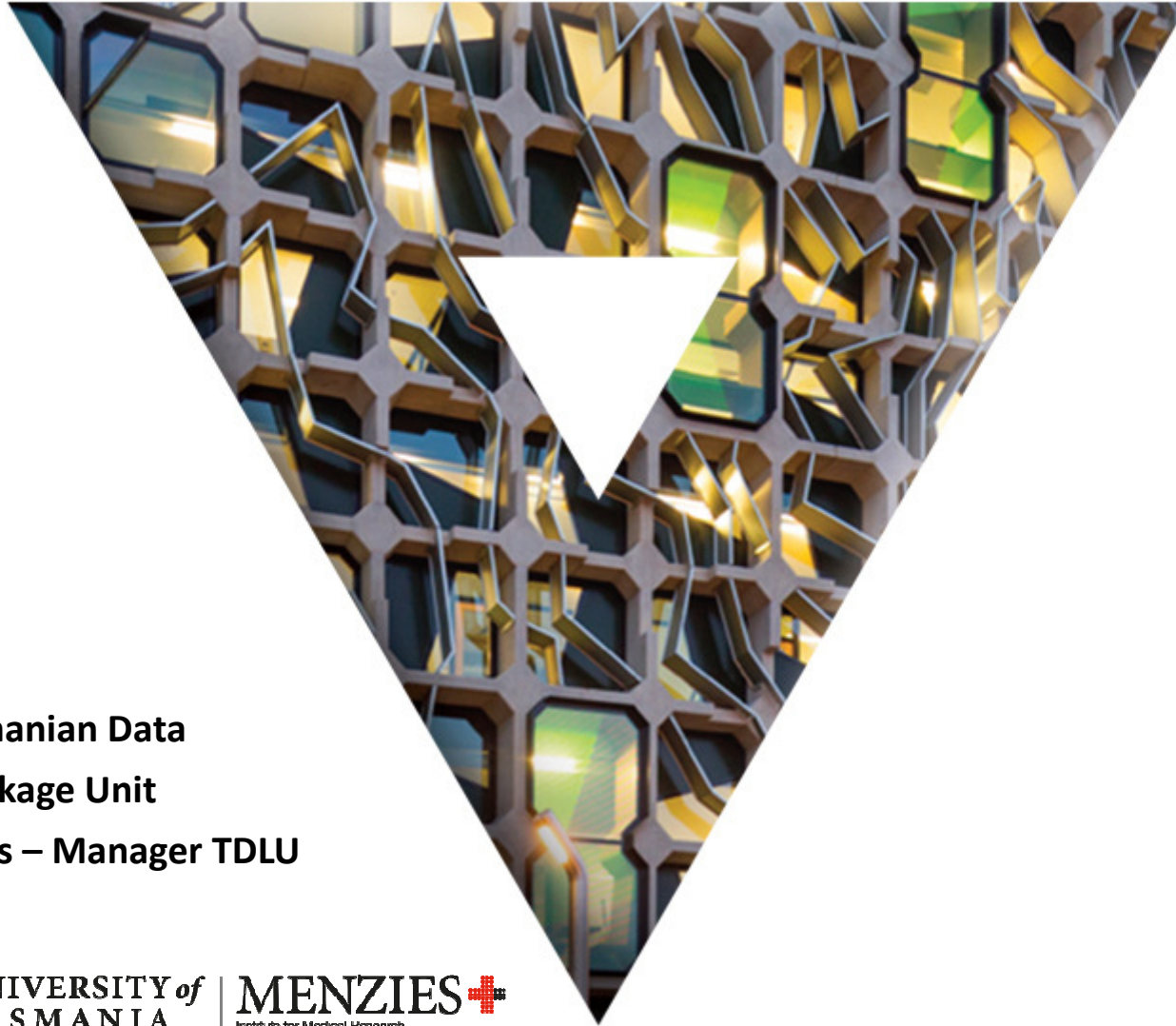


# Data Linkage in Tasmania



**Tasmanian Data**

**Linkage Unit**

**Brian Stokes – Manager TDLU**

# Overview of TDLU

- Established in 2010
- Current staff of 3, casual staff to undertake clerical review
- Commonwealth funded through the National Collaborative Research Infrastructure Strategy (NCRIS)
- Agreement to operate in Tasmania is between the Dept. of Health & Human Services Tasmania and University of WA
- TDLU operated by the Menzies Institute for Medical Research
- Part of the Population Health Research Network (PHRN)
- Formal agreements in place with custodians for supply of core datasets for linkage

# What is Data Linkage?

- Data linkage is a method of bringing together information about people, places and events in a way that protects individual privacy
- Data linkage brings together this information from disparate information sources
- This enables the construction of chronological sequences of events for individuals – often referred to as “cradle to the grave”
- In combination, these individual ‘chapters’ create a bigger story about the health and wellbeing of people in the community
- In bringing records together, the TDLU uses strict privacy preserving policies, protocols and procedures to ensure the security of the data and confidentiality of the individuals the records relate to.

# Model of Separation - 3 actors, privacy focus

- **Data Custodians**
  - Hold a full data set, administrative information plus clinical, service, activity data
  - Name, address, date of birth, date of death, sex, system identifier plus, for example, specific health, education, judicial or death data
  - I.e., ambulance events, school enrolments, hospital admissions, birth and death registrations, cancer records
- **Data Linkage Unit**
  - Obtain administrative variables only from custodians
  - Sufficient enough to enable linkage across disparate data sources
- **Researcher / End user**
  - Obtain individual sets of data from each custodian
  - De-identified and anonymised
  - Combine into a single researchable dataset using a unique key generated by the linkage unit

# Linkage Model: Privacy Focus

## Why the Separation Principle?

Preserves privacy

1. Custodians have access to their full data set
2. The linkage unit doesn't receive service, clinical or activity data
3. Researchers don't receive demographic data

Data release directly controlled by custodians

Data users sign binding agreements related to confidentiality and security

# Key Terms

- **Master Linkage Map (MLM)** - The MLM groups together records for individuals in a population. Each individual within the *Map* has their own unique 'key'.
- **Master Linkage Key (MLK)** – An individual's unique ID otherwise known as a 'key'; 1 MLM = Many Keys
- **Project Person ID (PPID)** - A project-specific *pseudo-identifier* that is supplied to data custodians that refer to individuals, and with minimal risk of re-identification
- **Source Identifier** – The unique identifier from the source dataset. For example, in the case of an admitted patient dataset, the Source Identifier will be the patient URN (unit Record Number)

# Distinct Parts of the Linkage Process

- From the administrative variables supplied to the TDLU, a process of 'Probabilistic Linkage' is undertaken
- Using sophisticated linkage software, we determine the likelihood of a person in Dataset A being the same as in Dataset B.
- Links are generated and stored in a Master Linkage Map (MLM).
- For every project, a unique Project Person Identifier (PPID) is generated for every record in the dataset
- The TDLU sends linkage keys back to custodians – source system identifier plus the PPID
- Custodians in turn attach approved data items, strip their source system identifier and send the resultant anonymised dataset to the researcher.

# Matching Challenges

Dataset	First Name	Surname	DOB	Suburb	Postcode
Birth	Sam	McCarthy	01-Jan-1970	Hobart	7000
Perinatal	Samantha	MacCarthy	01/01/70	Hobart	7000
Hospital Episode	Samantha	McCarthey	1/1/70	Hobart, 7000	
Cancer	Samantha	Jones	01/01/70	West Hobart	7004
Education	Samantha	MacCarthy	01/01/70	Hobart	7000
Criminal Justice	Samantha	Jones	01/01/70	Hobart	7004



# Matching Challenges

Dataset	First Name	Surname	DOB	Suburb	Postcode
Birth	Sam	McCarthy	01-Jan-1970	Hobart	7000
Perinatal	Samantha	MacCarthy	10/01/70	Hobart	7000
Hospital Episode	Samantha	McCarthey	1/10/70	Hobart, 7000	
Cancer	Samantha	Jones	01/10/70	West Hobart	7004
Education	Samantha	MacCarthy	01/10/70	Hobart	7000
Criminal Justice	Samantha	Jones	01/10/70	Hobart	7004

- 3 different spellings / variations of surname
- 2 variations of Christian name
- Possible change of surname – marriage etc.,
- Possible change of address
- Possible date transcription
- 1 possible incorrect postcode

# Matching Challenges

Dataset	First Name	Surname	DOB	Suburb	Postcode	MLK
Birth	Sam	McCarthy	01-Jan-1970	Hobart	7000	190876
Perinatal	Samantha	MacCarthy	10/01/70	Hobart	7000	190876
Hospital Episode	Samantha	McCarthey	1/10/70	Hobart, 7000		190876
Cancer	Samantha	Jones	01/10/70	West Hobart	7004	190876
Education	Samantha	MacCarthy	01/10/70	Hobart	7000	190876
Criminal Justice	Samantha	Jones	01/10/70	Hobart	7004	190876

- After clerically reviewing the groups, it was concluded that all records relate to the same individual
- Allocated the same MLK

# Master Linkage Map

Dataset	Total Records	Unique keys	Date From	Date To
Admitted patient episodes	910,258	254,307	1/7/2007	31/12/2014
Emergency Department Presentations	1,753,672	470,336	1/1/2000	31/12/2014
Registrar of deaths	170,405	170,180	1/1/1970	29/8/2015
Registrar of births	89,824	89,814	1/1/2000	31/12/2014
Perinatal – mother (public)	47,732	30,947	1/1/2005	31/12/2014
Perinatal – baby (public)	47,732	47,722	1/1/2005	31/12/2014
Perinatal – mother (pvte)	13,901	9,578	1/1/2005	31/12/2014
Perinatal – baby (pvte)	13,901		1/1/2005	31/12/2014
^ Cancer	83,903	74,060	1/1/1982	31/12/2012
^ Dept. of Education Census	385,177	84,076	2008	2010
Australian Early Development Census	557,518	556,765	2009 and 2012 collections	
^ Ambulance Tasmania – Cardiac Events	4,212	4,166	1/1/2007	31/12/2014
^ ad-hoc linkage				

# Master Linkage Map – Other datasets

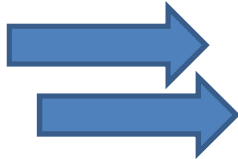
Dataset	Total Records	Unique keys
Strength to Strength	202	202
Care Coordination	3,060	2,918
ANZICS Core NW	1,085	990
ANZICS Core LGH	2,384	2,098
ANZICS Core RHH	5,115	4,256
<b>Project 1 - Chlamydia</b>		
<b>Public Hospital Anatomical Pathology</b>	<b>9,364</b>	<b>Total Persons</b> <b>22,173</b>
<b>Private Pathology</b>	<b>24,505</b>	

# Master Linkage Map – Other datasets

Dataset	Total Records	Unique keys
Strength to Strength	202	202
Care Coordination	3,060	2,918
ANZICS Core NW	1,085	990
ANZICS Core LGH	2,384	2,098
ANZICS Core RHH	5,115	4,256
<b>Project 1 - Chlamydia</b>		
Public Hospital Anatomical Pathology	9,364	Total Persons 22,173
Private Pathology	24,505	
<b>Project 2 – Thyroid Function testing (1994-2015)</b>		
Public Hospital Anatomical Pathology	146,396	Total Persons 430,467
Private Pathology	1,652,501	

# Master Linkage Map – Datasets to be added

Dataset	Date From	Date To
Admitted patient episodes	1/1/2015	31/12/2015
Emergency Department Presentations	1/1/2015	31/12/2015
Registrar of deaths	1/9/2015	30/4/2016
Registrar of births	1/1/2015	31/12/2015
Community Mental Health	1/1/2000	31/12/2015
Australian Early Development Census	2015 Collection	
Perinatal - Mother	1/1/2015	31/12/2015
Perinatal - Baby	1/1/2015	31/12/2015



# Links Made

Dataset	No other	AEDC	Amb	ANZIC LGH	ANZICS NWR	ANZICS RHH	AP	Births	Cancer	Census	Deaths	ED	Per Baby	Per Mother
AEDC	542,673	556,765	8				3,888	11,602	13	5,670	5	10,116	4,431	
Amb	550	8	4,166	103	52	202	2,431	58	712	92	3,281	3,301	42	39
ANZICS LGH			103	2,098	44	94	2,094		703	24	992	2,035		60
ANZICS NWR	2		52	44	993	71	990		257	19	394	968		32
ANZICS RHH	24		202	94	71	4,581	4,526		1,090	55	1,521	4,127		88
AP	15,939	3,888	2,431	2,094	990	4,526	254,307	44,434	25,131	23,273	28,094	213,883	36,256	26,193
Births	9,855	11,602	58				44,434	89,814	85	24,208	428	55,770	46,293	1
Cancer	10,573	13	712	703	257	1,090	25,131	85	74,060	283	45,126	34,392	37	320
Census	16,966	5,670	92	24	19	55	23,273	24,208	283	84,076	196	59,391	3,761	2,896
Deaths	92,188	5	3,281	992	394	1,521	28,094	428	45,126	196	170,180	50,220	240	86
ED	171,360	10,116	3,301	2,035	968	4,127	213,883	55,770	34,392	59,391	50,220	470,336	29,713	23,904
Per Baby	400	4,431	42				36,256	46,293	37	3,761	240	29,713	47,722	
Per Mother	1,647		39	60	32	88	26,193	1	320	2,896	86	23,904		30,947

# Linkage is possible with Commonwealth datasets

- Popular commonwealth datasets
  - National Death Index - NDI
  - Australian Cancer Database - ACD
  - Pharmaceutical Benefits Scheme - PBS
  - Medicare Benefits Scheme – MBS
  - Australian Childhood Immunisation Register – ACIR
- Australian Coordinating Registry – ACR



# Projects Completed – In Brief

Project	Principal Research Organisation	Linked Datasets
Standardised Hospital Mortality Ratio	DHHS	<ul style="list-style-type: none"> <li>- Admitted Patients</li> <li>- Tasmanian Cancer Registry</li> <li>- Deaths</li> </ul>
Factors Impeding Access to Defibrillation	University of NSW / Ambulance Tasmania	<ul style="list-style-type: none"> <li>- VACIS/CAD</li> <li>- Admitted Patients</li> <li>- Emergency Presentations</li> <li>- Deaths</li> </ul>
Survival of Tasmanian ICU Patients After Discharge from Hospital	ANZICS	<ul style="list-style-type: none"> <li>- Admitted Patients</li> <li>- Deaths</li> <li>- LGH, NWRH and RHH APD Datasets</li> </ul>
Care Coordination and Strength to Strength	Primary Health Tasmania	<ul style="list-style-type: none"> <li>- Care Coordination</li> <li>- Strength to Strength</li> <li>- Admitted Patients</li> <li>- Emergency Presentations</li> </ul>
Thyroid Monitoring	DHHS	<ul style="list-style-type: none"> <li>- Anatomical Pathology</li> <li>- Diagnostic Services Pathology</li> </ul>
Population Level Chlamydia Testing and Positivity Rates in Tasmania	Menzies Institute for Medical Research	<ul style="list-style-type: none"> <li>- Anatomical Pathology</li> <li>- Diagnostic Services Pathology</li> </ul>

# Projects Underway – In Brief

Project	Principal Research Organisation	Linked Datasets
Making it count: Conception to Community	Menzies Institute for Medical Research	<ul style="list-style-type: none"> <li>- Perinatal Baby and Mother</li> <li>- Admitted Patients</li> <li>- Emergency Presentations</li> </ul>
*Quantifying the Burden of systemic Sclerosis in Australia	St Vincent's Hospital Melbourne	<ul style="list-style-type: none"> <li>- Australian Scleroderma Cohort</li> <li>- Admitted Patients</li> <li>- Emergency Presentations</li> <li>- Tasmanian Cancer Registry</li> <li>- Medicare Benefits Scheme</li> </ul>
*Living with acute coronary syndrome	The George Institute for Global Health	<ul style="list-style-type: none"> <li>- ACS SNAPSHOT Cohort</li> <li>- Admitted Patients</li> <li>- Emergency Presentations</li> </ul>
*Birthplace in Australia	University of Technology Sydney	<ul style="list-style-type: none"> <li>- Perinatal Baby and Mother</li> <li>- Admitted Patients</li> <li>- Deaths</li> </ul>
*The burden and cost of injury attributable health care use and mortality in Australia	Macquarie University	<ul style="list-style-type: none"> <li>- Admitted Patients</li> <li>- Emergency Presentations</li> <li>- Deaths</li> <li>- Electoral Roll</li> </ul>
*Developmental trajectories in Australia: Perinatal outcomes and child development (risk and protective factors)	Telethon Kids Institute	<ul style="list-style-type: none"> <li>- Perinatal Baby and Mother Births and Deaths</li> <li>- AEDC, NAPLAN, and various State Education Datasets</li> </ul>

\* Cross Jurisdictional or Multi Jurisdictional Projects