



Using Data Linkage

James Harrison

Director, Research Centre for Injury Studies, Flinders University
Chair, Steering Committee SA NT DataLink



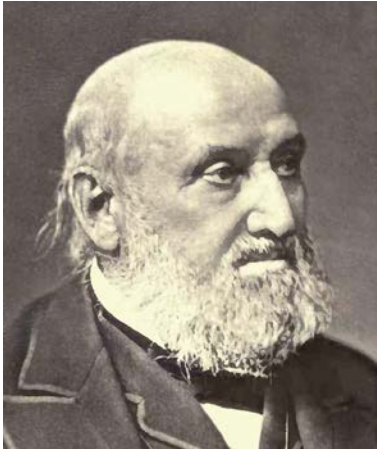
Acknowledgments

Andrew Stanley, Director SA NT DataLink

Chris Radbone, Associate Director, SA NT DataLink.

Population data to guide health care, policy and programs

Population data to guide health care, policy and programs



SIR,

London, 6th May 1839.

I beg leave to submit to you abstracts of the recorded causes of 141,607 deaths, which were registered in England and Wales during the half-year ending December 31st, 1837.

The registration of births and deaths proves the connection of families, facilitates the legal distribution of property, and answers several other public purposes which sufficiently establish its utility; but in the performance of the duty with which you have been pleased to intrust me, I have to examine the registration under a different point of view, and with different objects, which will perhaps ultimately prove of not less importance. The deaths and causes of death are scientific facts which admit of numerical analysis; and science has nothing to offer more inviting in speculation than the laws of vitality, the variations of those laws in the two sexes at different ages, and the influence of civilization, occupation, locality, seasons, and other physical agencies, either in generating diseases and inducing death, or in improving the public health.

Dr William Farr's opening words in the first report of the Registrar-General of the death registration system:

“... I have to examine the registration under a different point of view... deaths and causes of death admit of numerical analysis... variation of [mortality] in the two sexes at different ages and [under] the influence of civilization, locality, seasons, and other physical agencies, either in generating diseases and inducing death, or in improving the public health.”

6th May 1839

Population data to guide health care, policy and programs

The model of Farr & colleagues

- Death register *whole-population scope*
demographic data
causes of death
- Cause coding *developed into ICD*
- Denominators *census-based populations*
other (eg. person-miles by rail)
- Dissemination *statistical reports*
use for research projects

Population data to guide health care, policy and programs

A lot can be done with this model

- Infant mortality *Small-area rates*
Why shouldn't rates everywhere be as low as the lowest observed?
- Work-related *Miners' mortality high*
Risks greater & different in coal mining than other mining
- Suicide *Social risk factors*
Brought data to debate on suicide and educational attainment
- Data for others *Enabled Snow's work*

Population data to guide health care, policy and programs

Evolution of methods

- Mid 19th Century *Farr et al. model*
Few bytes per record. Human computers.
- Mid 20th Century *Electronic computers*
Faster & more flexible. Still small records.
- Late 20th Century *Some linkage of data*
Mainly project specific.
- Early 21st Century *Widespread linkage of data*
Persisting linkage systems. Diverse sources.
Some larger records (e.g. images).

Population data to guide health care, policy and programs

Evolution of methods

- Mid 19th Century *Farr et al. model*
Few bytes per record. Human computers.
- Mid 20th Century *Electronic computers*
Faster & more flexible. Still small records.
- Late 20th Century *Some linkage of data*
Mainly project specific.
- Early 21st Century *Widespread linkage of data*
Persisting linkage systems. Diverse sources.
Some larger records (e.g. images).

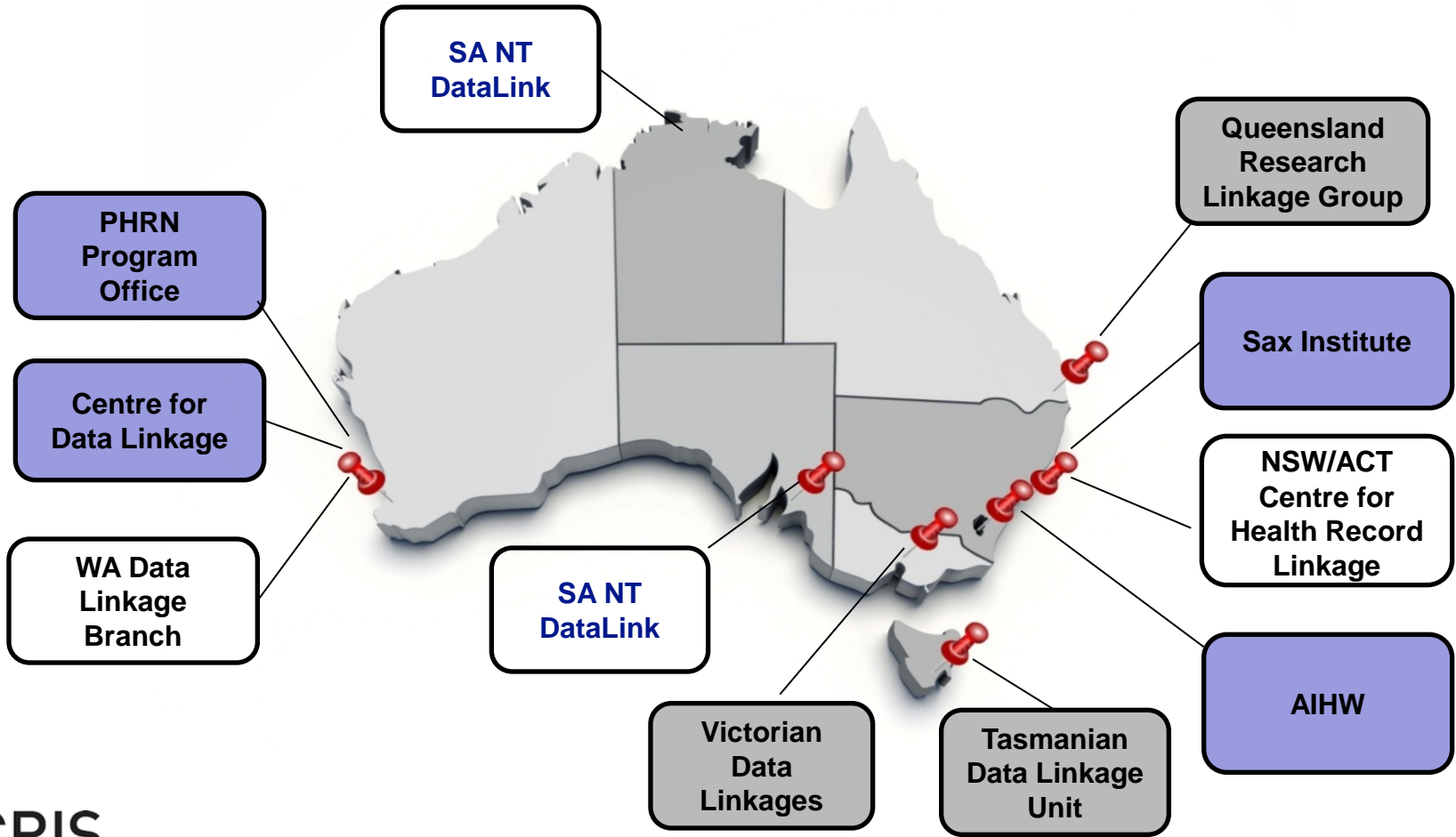
Population data to guide health care, policy and programs

Evolution of methods

- Mid 19th Century *Farr et al. model*
Few bytes per record. Human computers.
- Mid 20th Century *Electronic computers*
Faster & more flexible. Still small records.
- Late 20th Century *Some linkage of data*
Mainly project specific.
- Early 21st Century *Widespread linkage of data*
Persisting linkage systems. Diverse sources.
Some larger records (e.g. images).

Status? Potential? What next?

Population Health Research Network



The Australian Government has provided financial support to the PHRN and its members through the National Collaborative Research Infrastructure Strategy (NCRIS).



Who is SA NT DataLink?



SA NT Data Linkage consortium is an unincorporated Joint Venture – legally established & administered by the University of South Australia.

What does SA NT DataLink do?

What does SA NT DataLink do?

Has built and maintains a Master Linkage File

This is a file that records the presence of records for the same individual in a variety of other files.

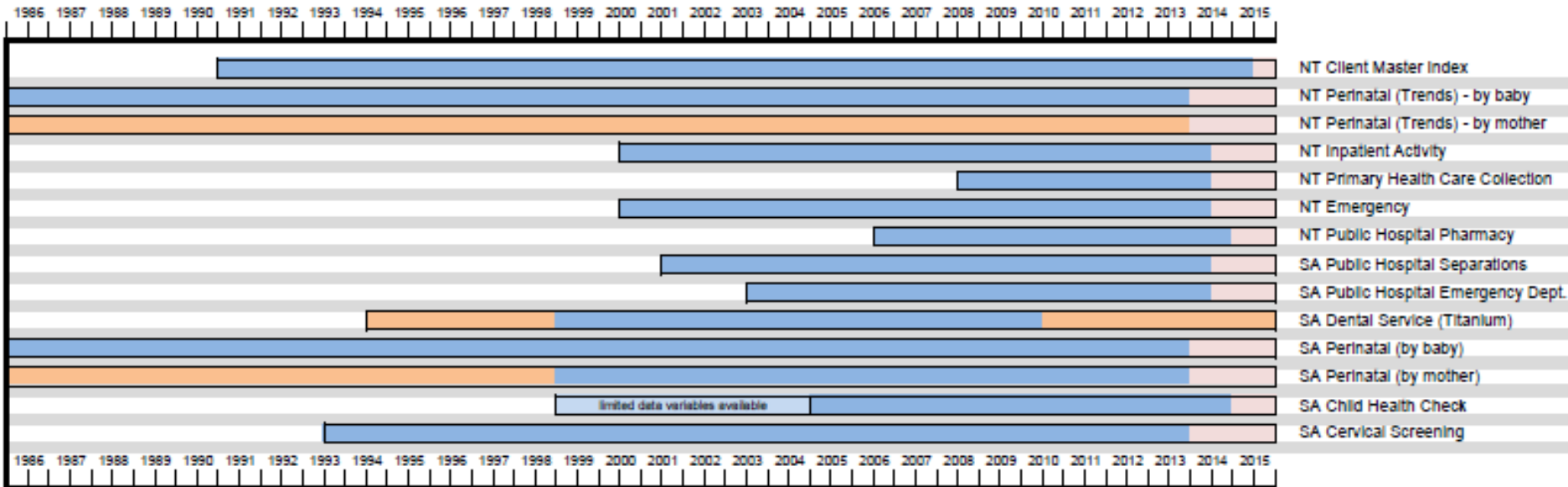
Making the master linkage file:

- requires identifying data (e.g. name)
- does not require 'content' data (e.g. which illness)

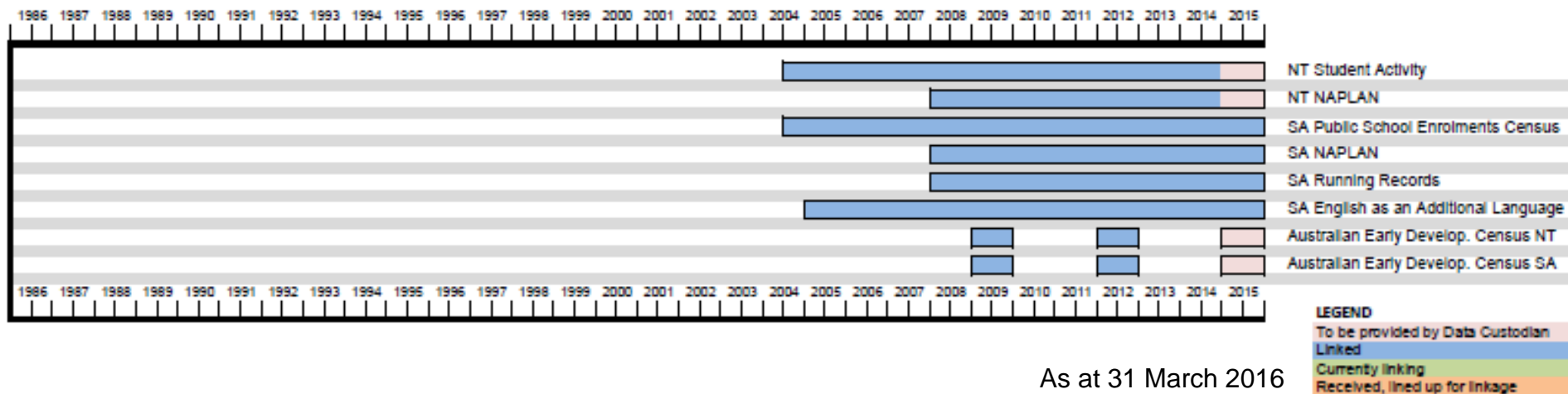
The Master Linkage file includes a large proportion of everyone in the SA and NT populations

What does SA NT DataLink do?

Health data



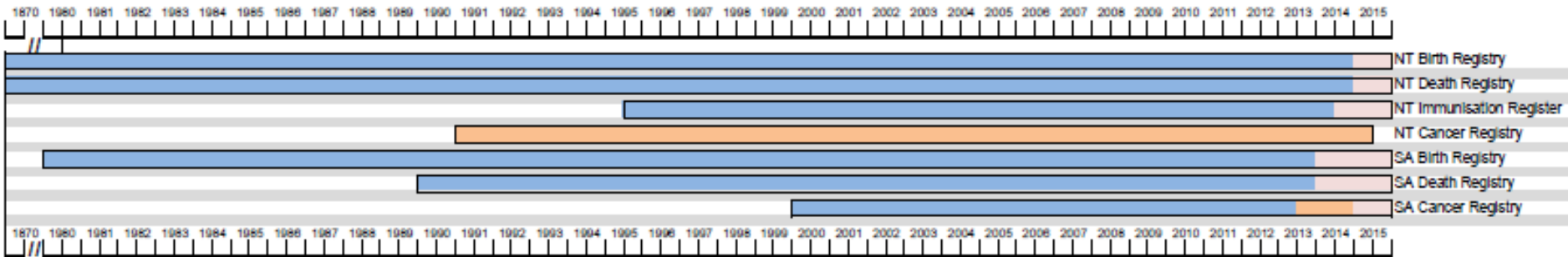
Education data



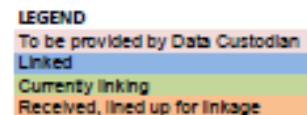
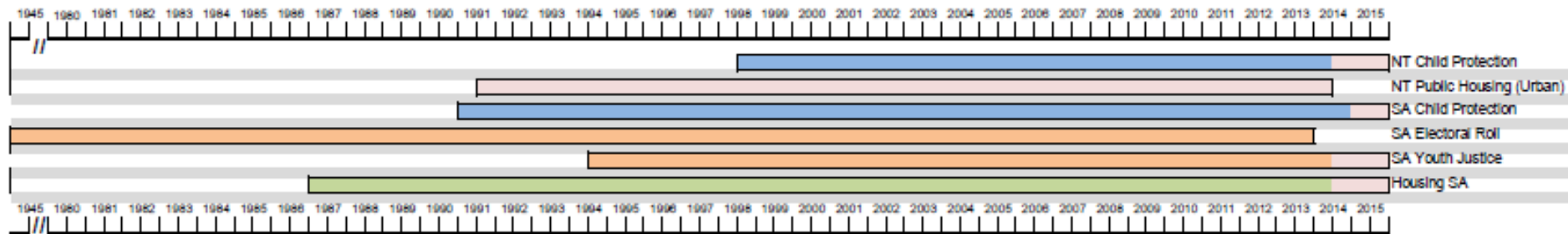
As at 31 March 2016

What does SA NT DataLink do?

Registries



Social data



As at 31 March 2016

What does SA NT DataLink do?

Has built and maintains a Master Linkage File

This is a file that records the presence of records for the same individual in a variety of other files.

Making the master linkage file:

- requires identifying data (e.g. name)
- does not require 'content' data (e.g. which illness)

The Master Linkage file includes a large proportion of everyone in the SA and NT populations

What does SA NT DataLink do?

Has built and maintains a Master Linkage File

This is a file that records the presence of records for the same individual in a variety of other files.

Making the master linkage file:

- requires identifying data (e.g. name)
- does not require 'content' data (e.g. which illness)

The Master Linkage file includes a large proportion of everyone in the SA and NT populations

Example of MLF content

What proportion of males in SA, who were born in particular years, are known to have at least one record in these data sources?

	Males born in		
	2006	1996	1945
Child Health Development Records	86%	0%	0%
Child Protection Data	20%	22%	0%
Australian Early Development Census	0%	0%	0%
Birth Registry	79%	0%	0%
Cancer Registry	0%	0%	20%
Death Registry	0%	0%	15%
Public School Student ID	59%	73%	2%
Public Hospital Emergency Department	72%	48%	72%
Public Hospital Inpatient Data	50%	28%	85%
Perinatal Records by Child	76%	0%	0%
Dental Records	7%	38%	20%

What does SA NT DataLink do?

Has built and maintains a Master Linkage File

This is a file that records the presence of records for the same individual in a variety of other files.

Making the master linkage file:

- requires identifying data (e.g. name)
- does not require 'content' data (e.g. which illness)

The Master Linkage file includes a large proportion of everyone in the SA and NT populations

What does SA NT DataLink do?

Has built and maintains a Master Linkage File

This is a file that records the presence of records for the same individual in a variety of other files.

Making the master linkage file:

- requires identifying data (e.g. name)
- does not require 'content' data (e.g. which illness)

The Master Linkage file includes a large proportion of everyone in the SA and NT populations

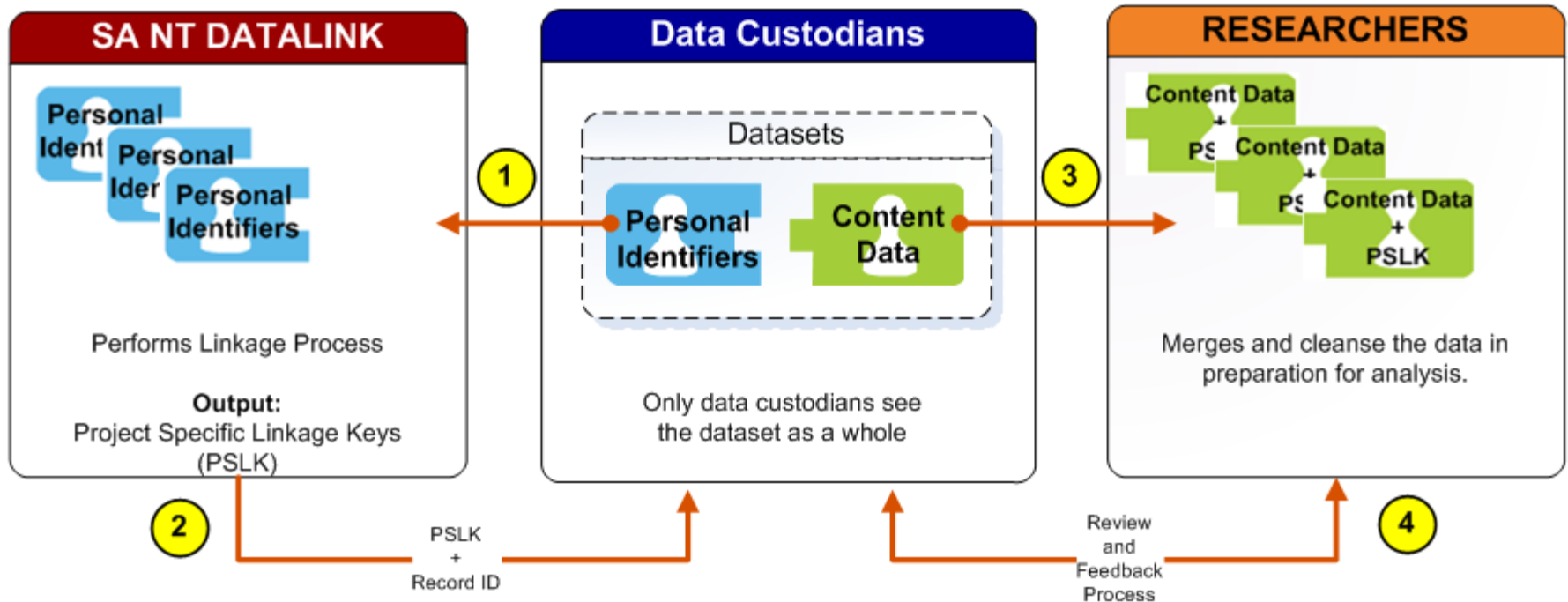
Facilitates research that makes use of the Master Linkage File

Once made, the master linkage file enables research using data on individuals without the researchers ever knowing their identities

- data custodians supply approved content data to researchers
- the supplied records specify individuals by means of a project-specific ID
- researchers use that ID to in place of names, etc. to join records on a person.

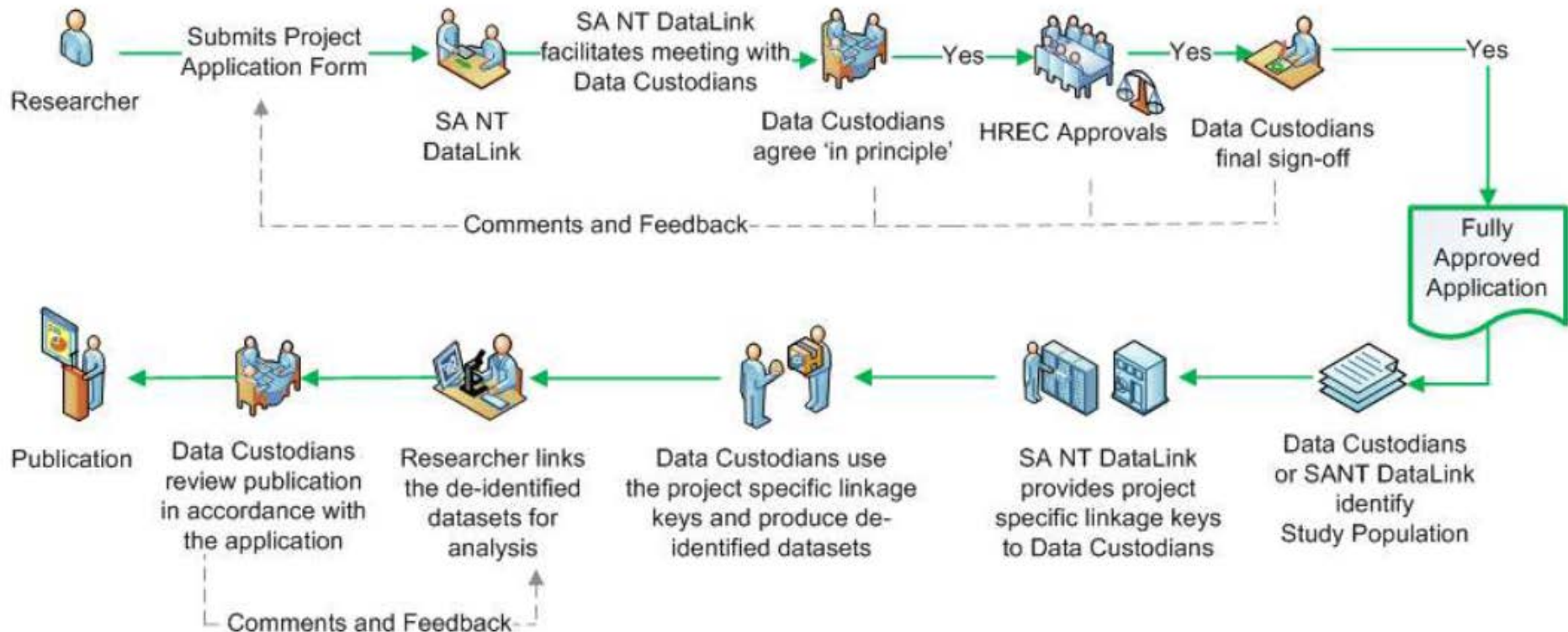
Security and Privacy Protection by Design

The Separation Principle



What does SA NT DataLink do?

Process for a project



What does SA NT DataLink do?

Examples of projects

- Early childhood development

Lynch & colleagues; many aspects

- Colorectal cancer

Roder & colleagues; Beckmann & colleagues.

Treatment and survival (whole population; over a long period)

- Injury risk factors and consequences

Mitchell SA data for national projects

Harrison TBI and school performance

What could SA NT DataLink do?

What could SA NT DataLink do?

Enable more research, and do that better (and faster)

Make once, use many times approach

Enable more than research ...

Providing an evidence base to monitor & improve the outcomes from government funded services



Commonwealth Government

trade, taxation, immigration, citizenship, social security, industrial relations and foreign affairs

State/Territory Governments

public health, education, roads, public land use, police, fire and ambulance services

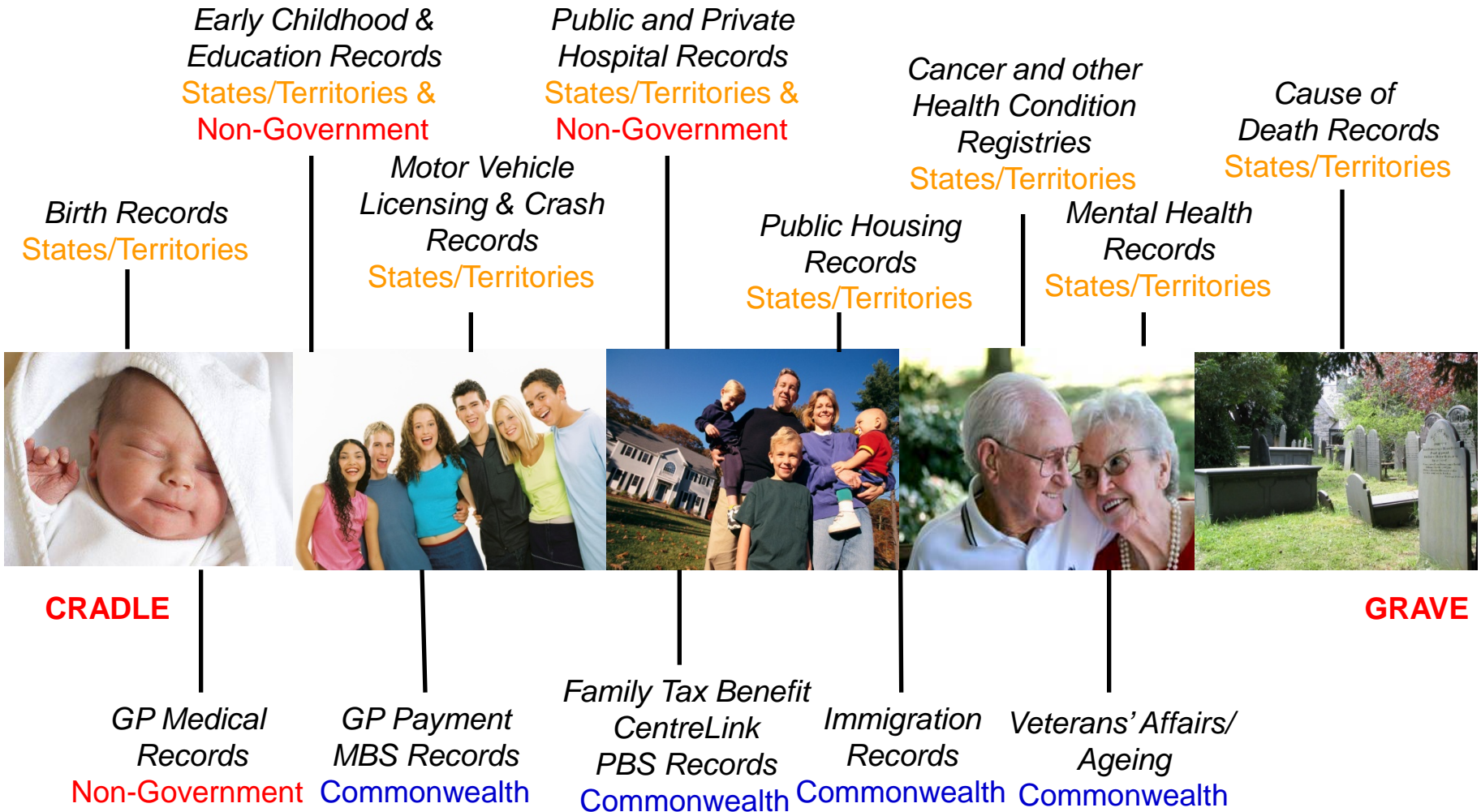
Joint Responsibility

Education & vocational training, transport, health and law enforcement

Local Government, Non-Government for Profit and Non-For-Profit sectors

Government funded service delivery

Tapping the potential of administrative data: Research, Government & the Community



Australian Government's Direction

December 2015 – Australian Government Public Data Policy Statement

Risk Management, IP & data ownership, Collaboration – Agencies, Jurisdictions
Standards – Interoperability; PM's 7 high priority innovation & data linkage projects

April 2014 - Better Practice Guide for Big Data including gaining **consent** for *future* analysis and use.

Underpinning Security to ensure the public's trust in the storage, access and approved use– **Protective Security Policy Framework & Information Security Manual**

August 2013 – “The Australian Public Service Big Data Strategy” AGIMO

Six principles to guide and assist Agencies:

1. Data is a national **asset**
2. **Privacy** by design
3. Data integrity (quality) and the transparency of processes
4. **Skills**, resources and capabilities will be shared
5. Collaboration with **industry and academia**
6. Enhancing **open data**



International Case Study Farr Institute (UK)

Worthwhile projects

Clear public benefit, scientifically and ethically sound

Data controllers opt in to each project

Safe people

Approved
researchers

Safe data

Limited
de-identified
data

Safe places

Secure data
centres

Safe outputs

SDC before
release of
results

Public engagement and communication

Tapping the potential of administrative data: Research, Government & the Community

Practical Challenges:

1. **Statistical issues** – linking data, and analysing the resulting linked datasets, raises a number of distinct challenges for researchers, although well-established methodologies and tools exist.
2. **Technical and operational issues** – gaining permission to access and use datasets held by multiple organisations may often be far from straightforward for researchers, and differences in the way data are collected may sometimes limit their use. Skills needed in Big Data Analysis.
3. **Institutional issues** – a range of legal, ethical and cultural considerations may significantly constrain the extent to which analyst can link data in practice. These may include variations and uncertainties over what is permissible, questions around consent, and concerns over public acceptability and trust.



Source: Enabling Data Linkage to Maximise the Value of Public Health Research Data: Full Report, The Wellcome Trust, UK, March 2015

Contacts

James Harrison
james.harrison@flinders.edu.au

Andrew Stanley, Director, SA NT DataLink
andrew.stanley@unisa.edu.au

Chris Radbone, Associate Director, SA NT DataLink
chris.radbone@unisa.edu.au

www.santdatalink.org.au

