

Victorian Data Linkages

HDSS Forum

3 June 2011

A/Prof Vijaya Sundararajan

Project Director

Victorian Data Linkages



Today

- Background and update on VDL
- Data linkage and research
- Privacy and data linkage
- Next steps

VDL Principles

- Data are a strategic resource for the public good.
- Data custodians, as they are accountable for their data's security and confidentiality, are at the centre of any data linkage.
- Data linkage should occur only in the public interest.
- Data linkage must be used for statistical and analytical purposes only.
- Policies and procedures supporting data linkage must minimise any potential impact on privacy and confidentiality.
- Data linkage will be conducted in an open and accountable way.

Funded by DBI and NCRIS(DIISR) to enhance state and national research infrastructure in Victoria

Objectives are to:

- ❖ Establish a map of population linkable data for Victoria
- ❖ Promote use of linked data among researchers
- ❖ Participate as the Victorian node in the national effort

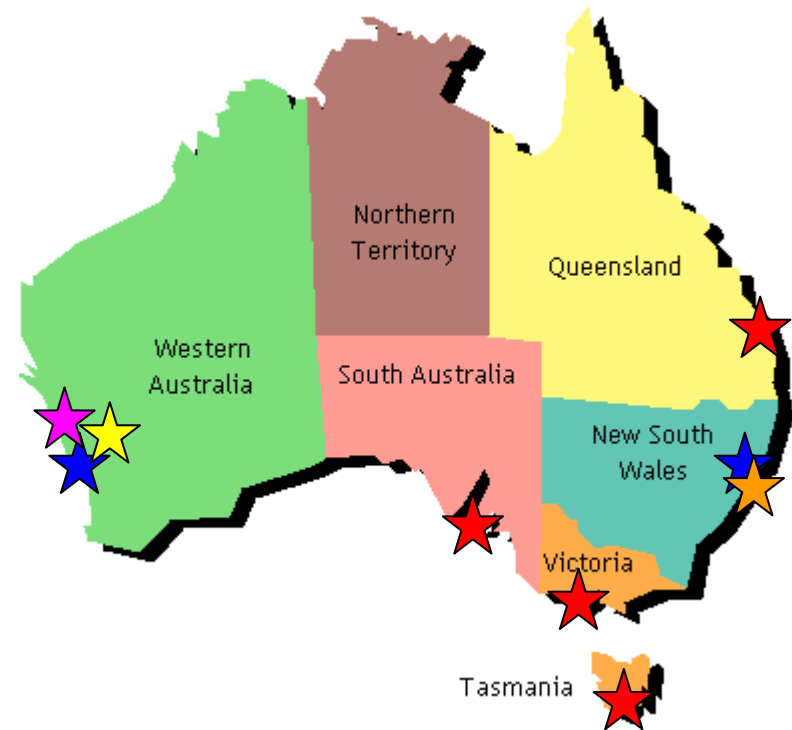
A national node *and* independent actor

Population Health Research Network (PHRN)

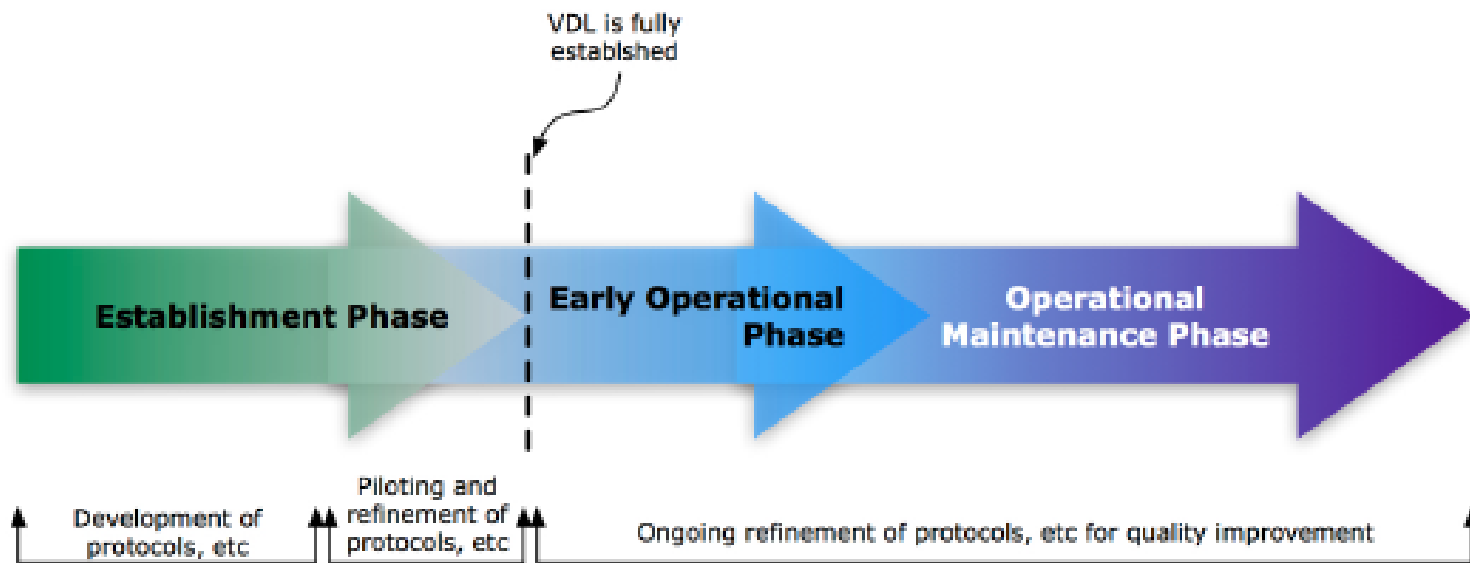
- ❖ PHRN is a new Network that is developing data linkage capability **within and between Australian states and territories**
- ❖ The PHRN has built on the experience of:
 - ❖ WA Data Linkage Unit established in 1994/95
 - ❖ The NSW Centre for Health Record Linkage (CHeReL) established in 2006
- ❖ It will implement a **common, best practice approach** to health data linkage
- ❖ This will **facilitate access** to linkable non-identified data for important population research

PHRN Structure

- ❖ Six state/territory **Data Linkage Units** including two existing units and four new units
- ❖ **National linkage centre**
- ❖ **Program Office** to providing coordination and client services
- ❖ A secure **Data Exchange/ Data Analysis Laboratory prototype**



Implementation of the VDL initiative



- Privacy policy, associated privacy framework and privacy management plan drafted.
- Research access: an application protocol being trialled to support researchers in their efforts to navigate the HREC, data custodians and associated privacy requirements.
- Website
- Lexicon to build a common language
- User-friendly data dictionaries

Victorian data linkages - Department of Health, Victoria, Australia - Windows Internet Explorer

http://www.health.vic.gov.au/vdl/

File Edit View Favorites Tools Help

Convert Select

Ask Search

CNN YouTube Facebook Games Celebrity Amazon Word of the Day Weather E-mail

Options

Victorian data linkages - Department of Health, Victori...

health Department of Health, Victoria, Australia

Email Print Search all of Health Search

Victorian Data Linkages

Menu

- Home
- For data custodians
- For researchers
- Staff overview
- Contact us

Health home > Victorian Data Linkages

About us

Victorian Data Linkages (VDL) has been established with funding from the Victorian Government Department of Business and Innovation and The Australian Government Department of Innovation, Industry, Science and Research National Collaborative Research Infrastructure Strategy (NCRIS) to develop new data linkage capacity in Victoria.

VDL is also a state node of the [Population Health Research Network \(PHRN\)](#).

The aims of VDL are to:

- Provide population wide linkable data to researchers while adhering to Victorian privacy principles;
- Build skills in and promote research using linked data in Victoria;
- Create opportunities for Victorian researchers to undertake innovative, world class research using linkable data drawn from existing datasets.

In order to achieve these goals, VDL will:

- Create a Victorian data linkage system focusing broadly on data collections that, when combined, will support research into health and wellbeing;
- Build a technology platform which ensures the accurate, efficient and secure creation of a linkage map between these datasets;
- Create tools to facilitate the use of linkable datasets by researchers;
- Develop privacy policies, protocols and procedures to ensure that the use of data by VDL and release of data to researchers adheres to health privacy principles; and
- Participate in the Population Health Research Network.

Anticipated outcomes from VDL include:

- Increased population health research achievements;
- Extension of national data linkage capacity through the Population Health Research Network; and
- Increased knowledge and skill in the research community in using linked data.

What's new

- ▶ Engaging with possibilities

Downloads

- ▶ [Introducing Victorian Data Linkages \(pdf, 2.37mb\)](#)

Done

Victorian data linkages - Department of Health, Victoria, Australia - Windows 1

Trusted sites

100%

Start

Vijaya Sundararajan - D...

J:\VDL\VDL Budget\EIF

Victorian data linkage...

M M

2:19 PM

Data Linkage

*“is a technique that makes possible
more complete use of health and other data
by bringing together data from many sources
at the level of
individual persons, populations, events or places,
which is often needed to
fully understand a population’s health”*

Friedman DJ, Hunter EL, Parrish RG. (2005).

Health statistics: shaping policy and practice to improve the population’s health.

New York: Oxford University Press

Data linkage and primary and secondary uses

‘Generally, when an agency collects information it does so for a specific purpose to meet an activity or function.

Data linkage occurs if this information about an individual, collected for one purpose, is combined with information or data about the same individual from another database, where it may have been collected for a different purpose.

This linked information can then be used for other purposes, perhaps unrelated to the original purpose of the collections, to create a new personal or health information “picture” or set of information about an individual.’



Data and researchers

‘researchers generally do not need to know the identity of the individual, simply that certain health information relates to the same individual.

This can be achieved through processes whereby **independent** intermediaries perform the linking of information, but do not have access to actual health information and researchers have access to the linked health information but not the identity of the individual’.

Australian Law Reform Commission, 2008.

For Your Information: Australian Privacy Law and Practice (Report 108),
Section 66 — Research: Databases and Data Linkage, emphasis added

What types of research?

Victorian Data Linkages supports research projects that are:

- in the **public interest** (i.e., for purpose of research rather than to support commercial activities);
- **approved** by a relevant Human Research **Ethics** Committee;
- **approved** by the relevant **data custodian(s)**;
- conducted **in accordance with the relevant legislation, regulations and guidelines**; and
- for which researchers undertake to **report the results in the public domain**.

The classic 'Kelman-Bass Best Practice Protocol'

- P Maximise the protection of individual **privacy**
- R Provide linked data files only to nominated
- I researchers involved in **specific, approved research projects**
- N E Provide researchers with **no more than the datasets**
- C S **required** for their specific project
- I Assure data custodians that those data which are
- P their responsibility will be **used appropriately and**
- L **that security obligations will be met.**

CHAPTER 3.2: DATABANKS

INTRODUCTION

This chapter covers a wide range of data types and methodologies. Given that the nature of data, data collection, research methodologies and data usage may change over time, the chapter presents principles rather than prescriptions.

Types of research that commonly make use of databanks include epidemiology, pathology, genetics and social sciences.

The term 'databanks', as used in this National Statement, includes databases.

What are data?

Data are pieces of information, for example:

- what people say in interviews, focus groups, questionnaires, personal histories and biographies;
- analysis of existing information (clinical, social, observational or other);
- information derived from human tissue such as blood, bone, muscle and urine.

Data identifiability

Data may be collected, stored or disclosed in three mutually exclusive forms:

- **individually identifiable data**, where the identity of a specific individual can reasonably be ascertained. Examples of identifiers include the individual's name, image, date of birth or address;
- **re-identifiable data**, from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets;
- **non-identifiable data**, which have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified. A subset of

non-identifiable data are those that can be linked with other data so it can be known that they are about the same data subject, although the person's identity remains unknown.

This National Statement avoids the term 'de-identified data', as its meaning is unclear. While it is sometimes used to refer to a record that cannot be linked to an individual ('non-identifiable'), it is also used to refer to a record in which identifying information has been removed but the means still exist to re-identify the individual. When the term 'de-identified data' is used, researchers and those reviewing research need to establish precisely which of these possible meanings is intended.

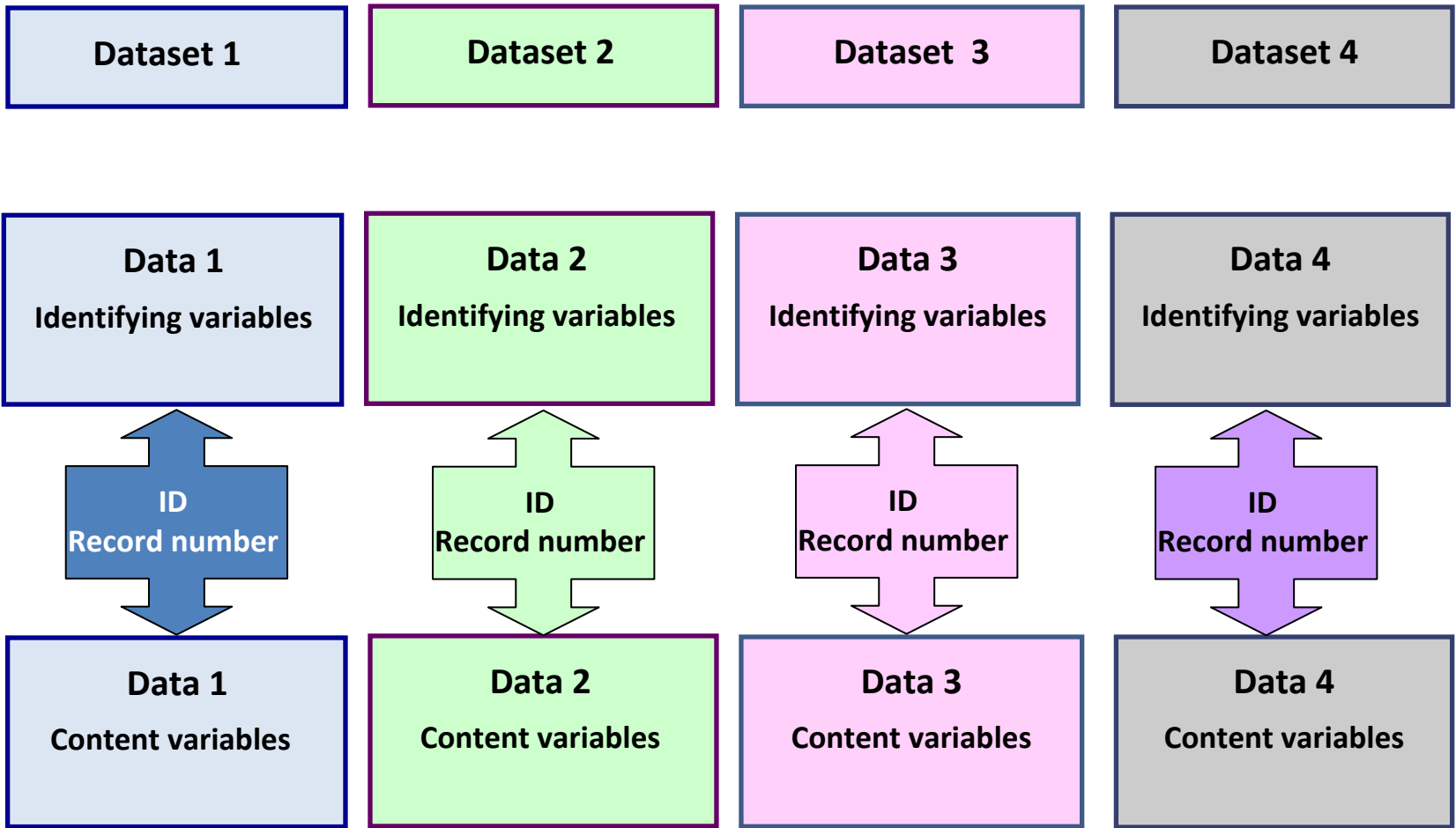
Tissue and data

With advances in genetic knowledge and data linkage, and the proliferation of tissue banks of identified material, human tissue samples should always be regarded as, in principle, re-identifiable.

The increased ability to link data has greatly enhanced the contribution that collections of data can make to research, as it enables researchers to match individuals in different data sets without being able to identify the person. For example, in epidemiological research (concerned with the study of populations), information about individuals and groups may be collected so that features of groups of people can be investigated. These data may or may not have originally been obtained for research purposes.

Banking

While most data are collected, aggregated and stored for a single purpose or activity. Permission may sometimes be sought from participants to 'bank' their data for possible use in future research projects.



Original data files from 4 different data custodians

Data 1

1. Replace Original IDs with encrypted Source IDs

2. Extract

- Data 1 Source IDs
- Data 1 Record No
- Linkage Variables

Data 2

1. Replace Original IDs with encrypted Source IDs

2. Extract

- Data 2 Source IDs
- Data 2 Record No
- Linkage Variables

Data 3

1. Replace Original IDs with encrypted Source IDs

2. Extract

- Data 3 Source IDs
- Data 3 Record No
- Linkage Variables

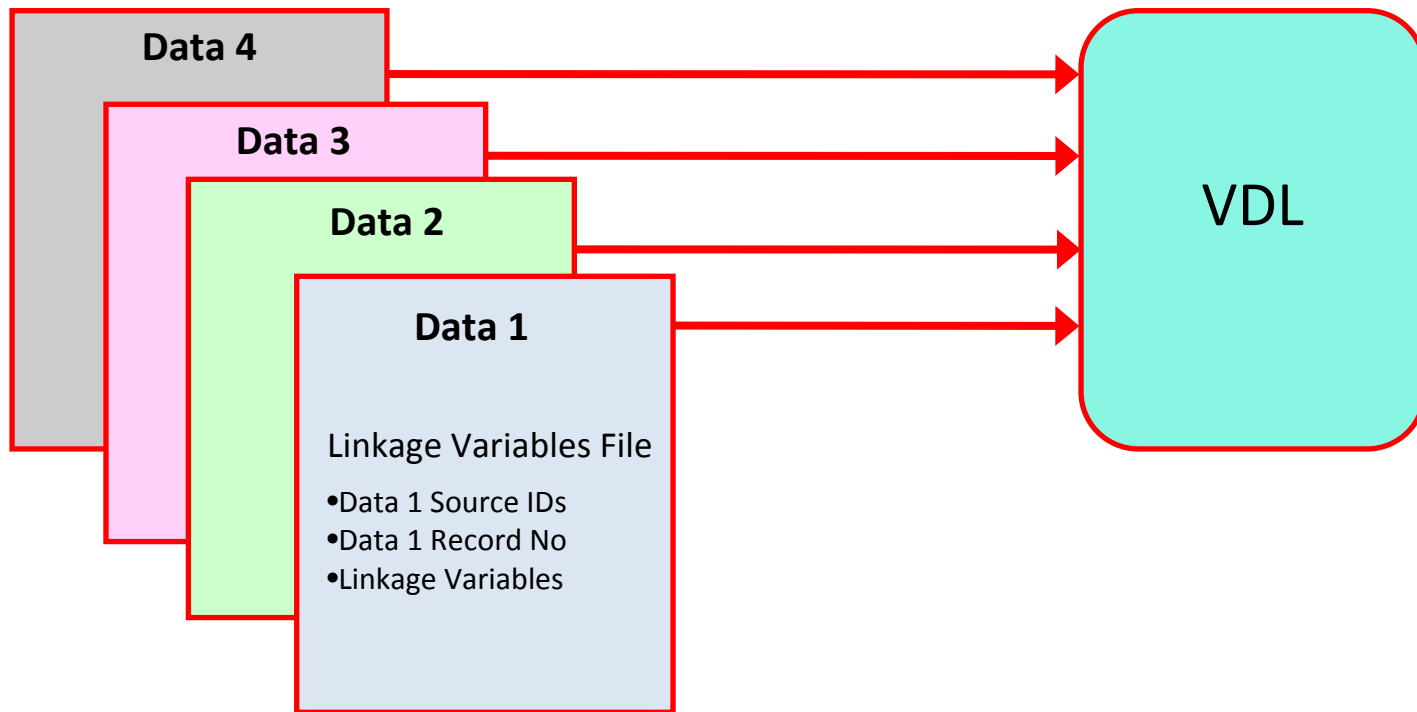
Data 4

1. Replace Original IDs with encrypted Source IDs

2. Extract

- Data 4 Source IDs
- Data 4 Record No
- Linkage Variables

STEP ONE – Data Custodians Prepare Linkage Variable Files

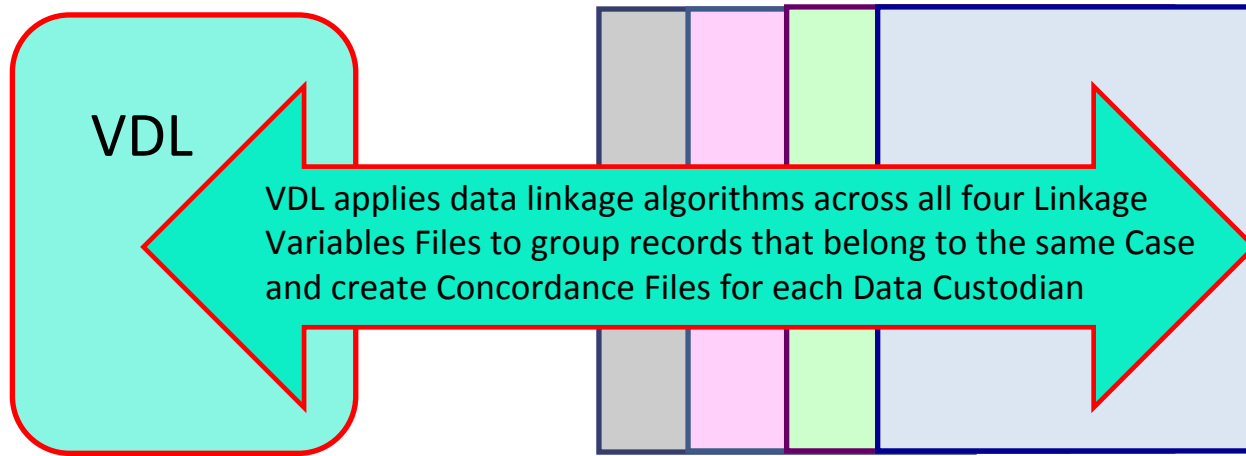


STEP TWO – Data Custodians Provide Linkage Variables Files to VDL

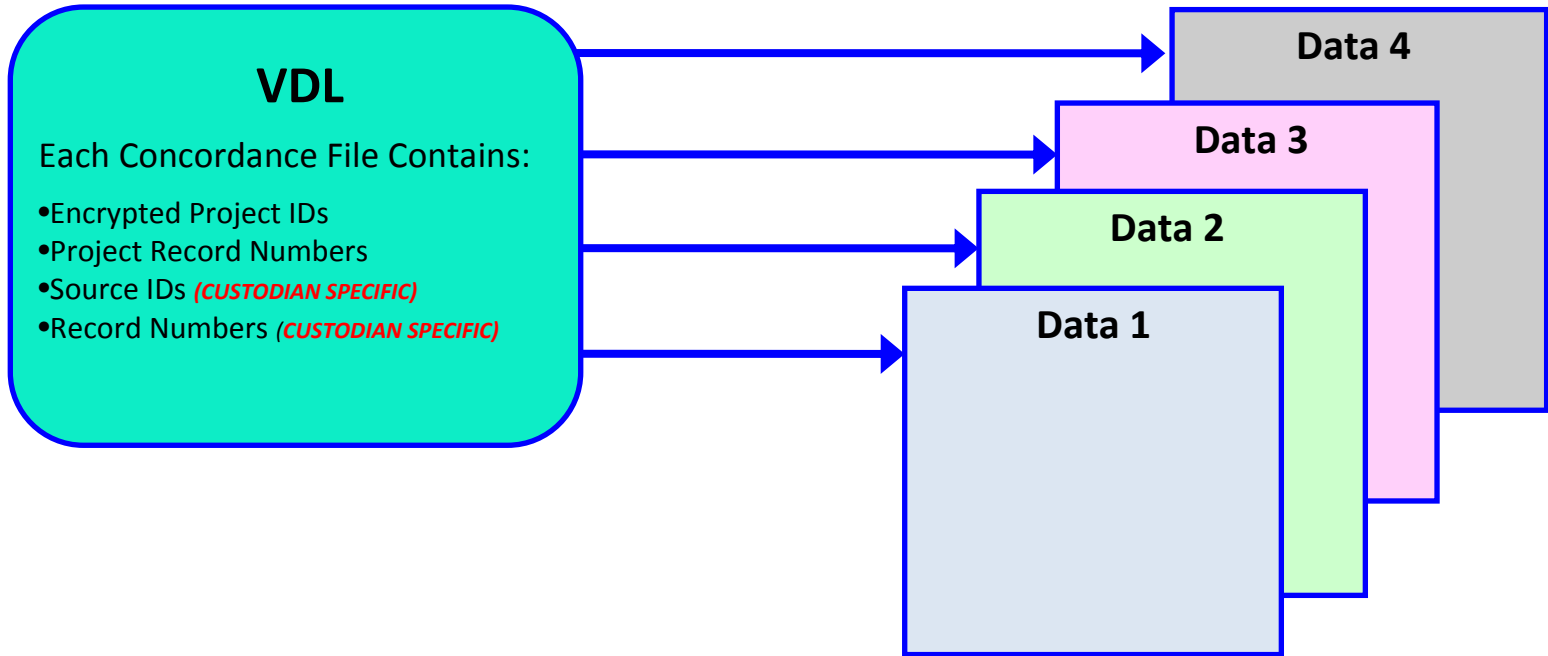


Identifying data

Individually identifiable data, where the identity of a specific individual can reasonably be ascertained. Examples of identifiers include the individual's name, image, date of birth or address.



STEP THREE – VDL Conducts Data Linkage Process Across Linkage Variables Files

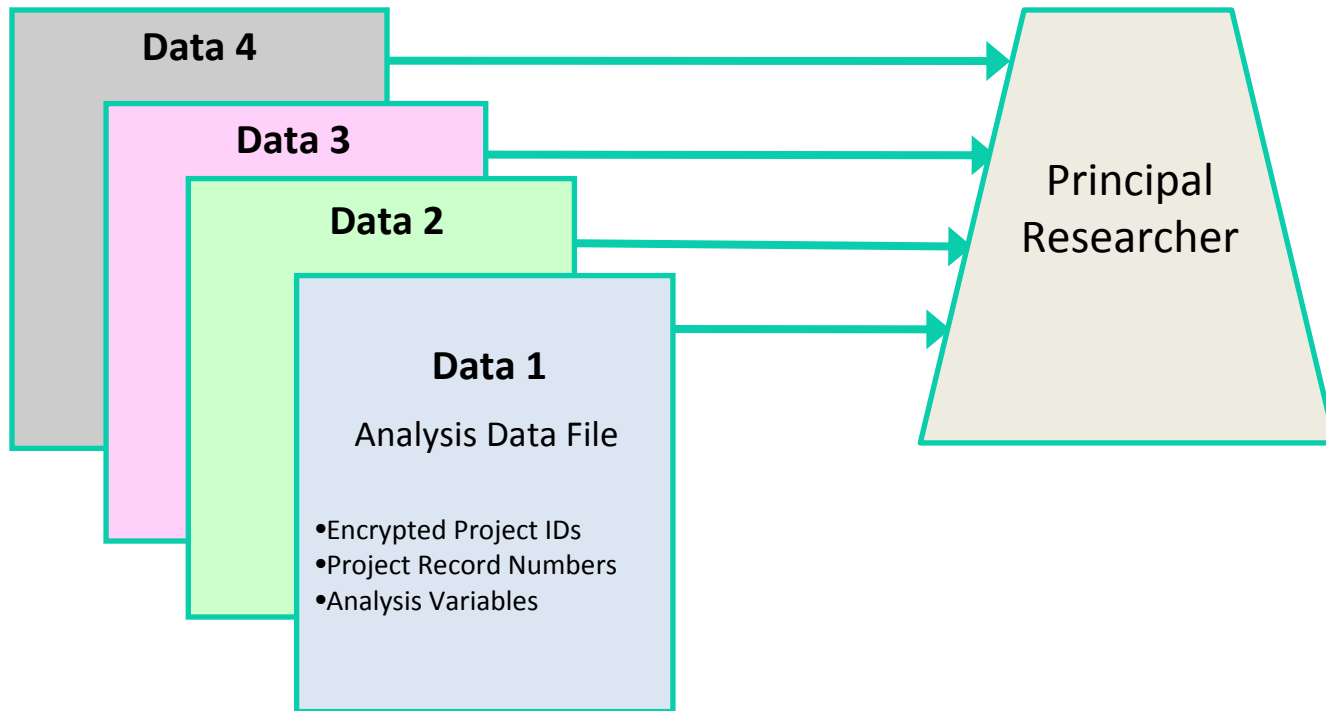


STEP FOUR – VDL Provides a Concordance File to Each Data Custodian



Re-identifiable data

Re-identifiable data, from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets.



STEP FIVE – Data Custodians Provide Analysis Data File to Principal Researcher

—————→
Non-identifiable data

Non-identifiable data, which have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified.

N
A
T
I
O
N
A
L

S
T
A
T
E
M

“A **subset non-identifiable data** are those that can be **linked with other data** so it can be known that they are about the same data subject, although the person’s identity remains unknown.”

The balance

Step 1: Separation of data items

Only direct identifiers or potentially identifying data items common to datasets are provided to data linkage unit by data custodians

Step 2: Production of linkage map

Linkage keys are created using identifying data items (only data linkage officers have access to the identifying data from the source datasets)

Step 3: Creation of linked data

Analysis files available to researcher include only the minimum clinical/service data items and demographic data items at the lowest level of identifiability consistent with research objectives, as approved by data custodians and HREC

limited risk to
protection of
privacy

reduced risk to
protection of
privacy



A period of moderate short-term risk
to achieve a long-term low risk dataset for the public good

1st research project for a linkage: "project specific linkage"

Datasets thought to have strong research interest and population coverage

Enduring, regularly updated linkage

ANZICS Registry
AUSCAR
MIG and Cardiac Surg
VACIS
SEHQ
Peri-natal
VICNISS
VCCR-BreastScreen
ASPREE
VCR

22

28

36

ific

Overarching 'parent' agreement (MOU) negotiated between VDL and data custodians including HREC for 'enduring linkage'

VAED

VEMD

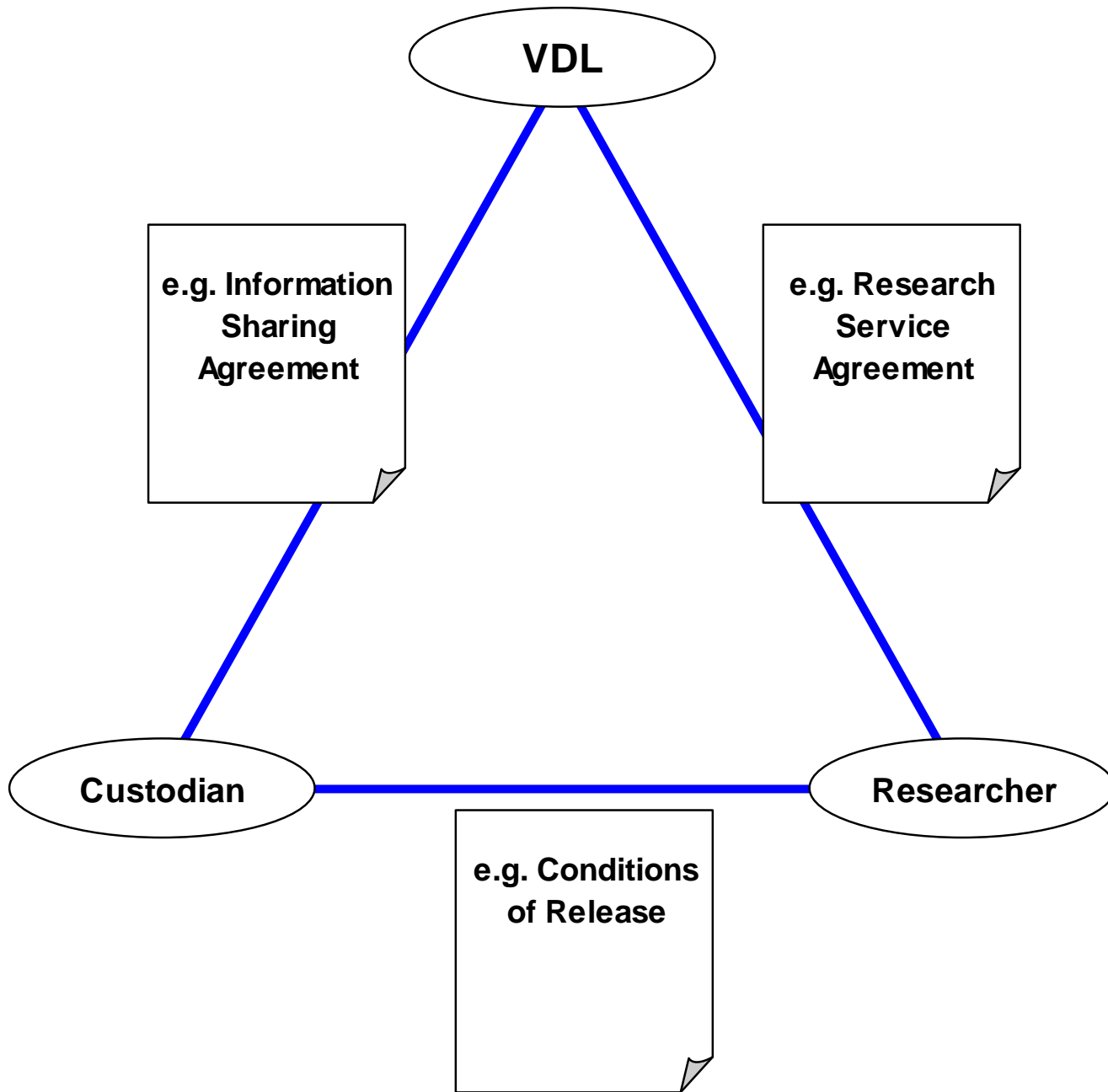
Death Registrations

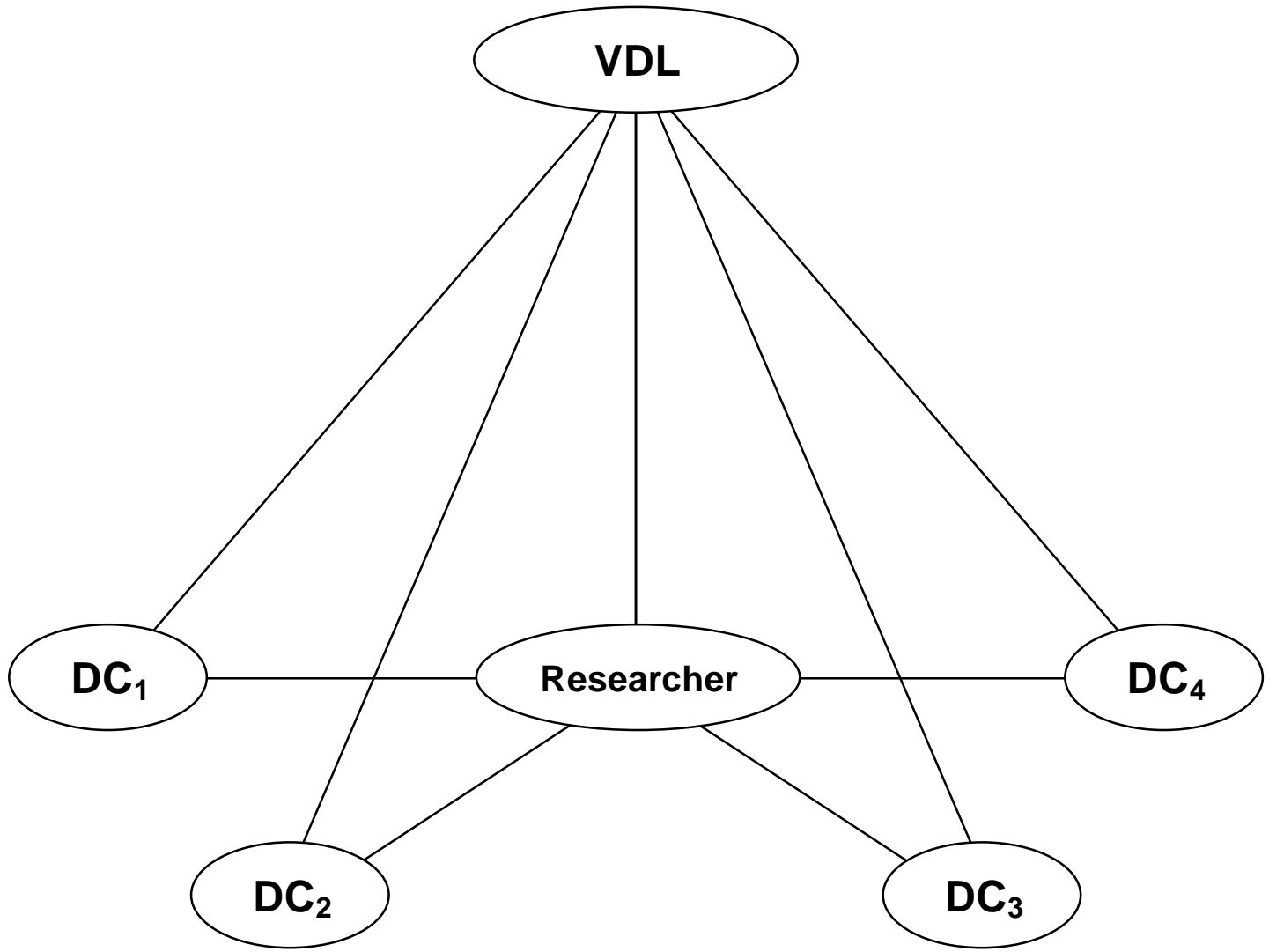
Master Linkage

Datasets	Finalised (to 2009-10)	Provisional (to Jan 2011)
VAED (from 1993-94)	29,537,980	30,596,321
VEMD (from 1999-00)	12,598,028	13,441,638
VDI (from 1994*)	577,926	-
*Under development		

Research data requests since 1 July 2010

Type	Count
Initial inquiry	7
IPD's being developed by researchers	14
IPD's submitted	17
IPD's reviewed	16
Datasets released (extract)	31
Datasets released (Linkage)	8
Dataset's re-released	2
Internal	9
External	32





www.health.vic.gov.au/vdl

