A Dynamic Data Linkage System – what's involved?

Carol Garfield [1]

Alfred John Bass [2]

Diana Rosman [1]

1 Department of Health, Perth, Western Australia
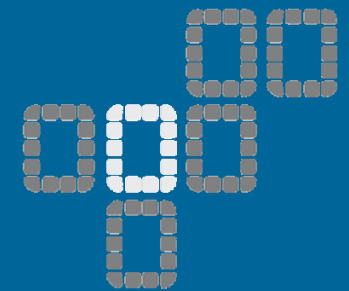2 Curtin University of Technology, Bentley, Western Australia
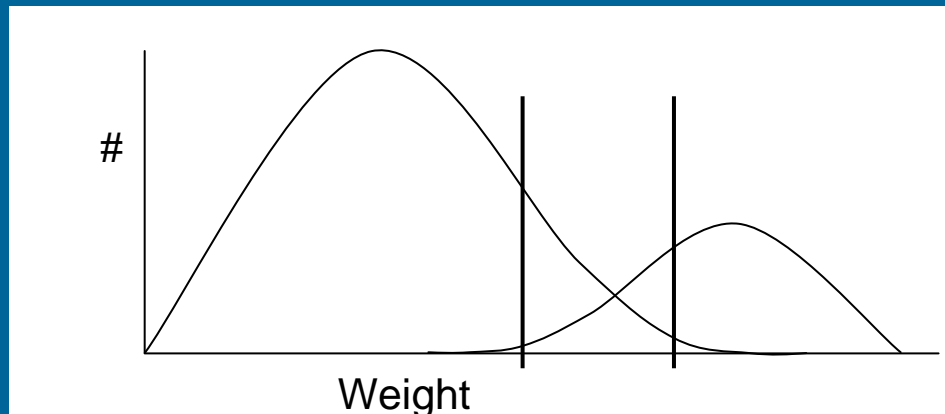
Objective

To describe the internal workings of the Western Australian Data Linkage System (WADLS)

➢ What is data linkage?

➢ WADLS:

➢ Creation, storage and extraction of links

➢ Current processes and future directions
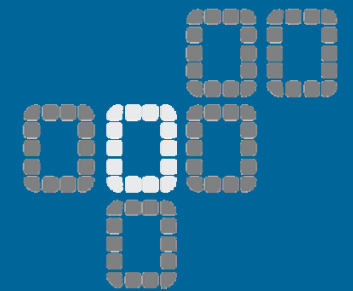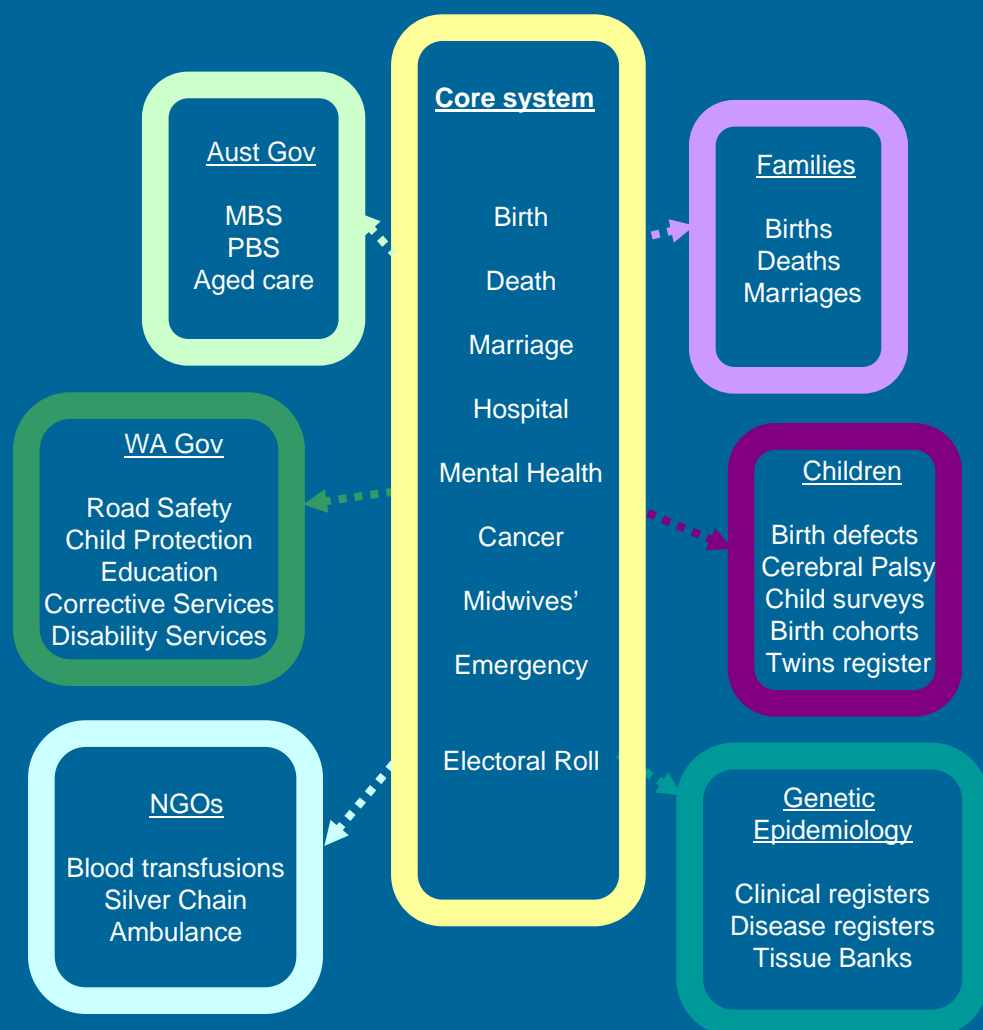
What is data linkage?

- a <u>technique</u> for creating links within and between data sources for information that is <u>thought to relate to</u> the <u>same</u> person, family, place or event

- probablistic methods compare like elements and calculate the likelihood that records belong to the same entity

- thresholds are set to separate probable and improbable links with those between requiring manual review

WA Data Linkage System 2008

Core system

- Birth
- Death
- Marriage
- Hospital
- Mental Health
- Cancer
- Midwives'
- Emergency
- Electoral Roll

Aust Gov

MBS
PBS
Aged care

WA Gov

Road Safety
Child Protection
Education
Corrective Services
Disability Services

NGOs

Blood transfusions
Silver Chain
Ambulance

Families

Births
Deaths
Marriages

Children

Birth defects
Cerebral Palsy
Child surveys
Birth cohorts
Twins register

Genetic Epidemiology

Clinical registers
Disease registers
Tissue Banks

Differs from a static linkage which is used to link 2 or more sources as a one off and for a specific purpose. For example, linkage of emergency department attendance to road crash events to measure injury severity and fatality rates

Multiple sources of data updated regularly (as often as monthly)
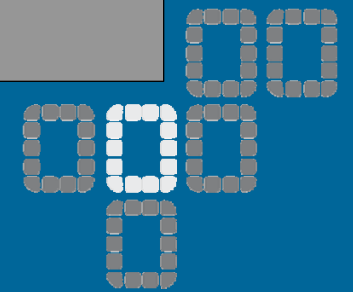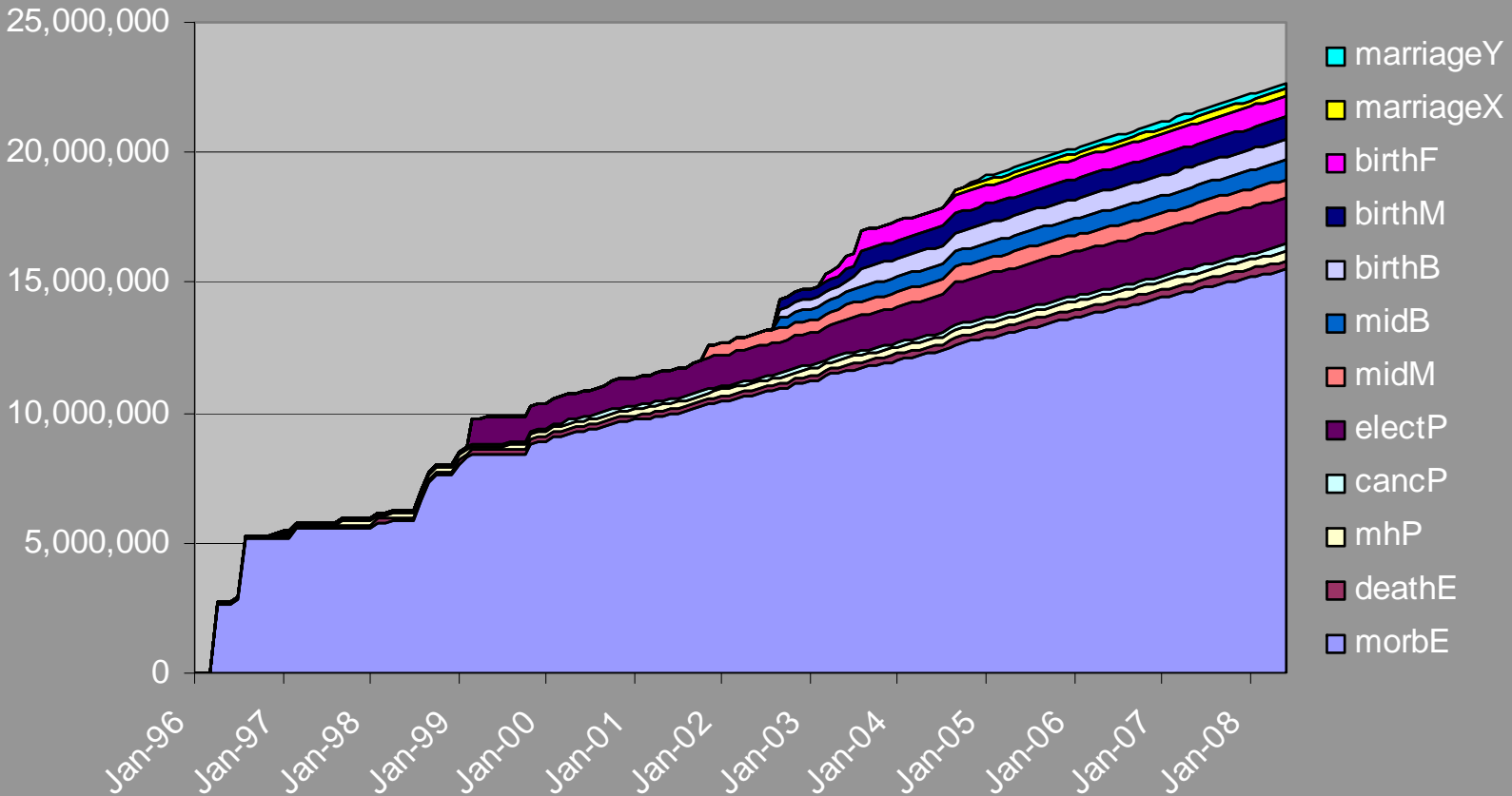- > 28 data sets
- > 38 millions records
- Records dated back to 1966

Dynamic with average of 10,000 links daily (for 2007)

Established as an infrastructure and used for research and planning

Volume of links for core datasets 1995-2008

Legend:
- marriageY
- marriageX
- birthF
- birthM
- birthB
- midB
- midM
- electP
- cancP
- mhP
- deathE
- morbE

Creation and Storage of Links

- Key design concept is the separation of linkage processes from the service or clinical data used for analysis
- WADLS uses demographics such as name, DOB, sex, address to create links using a commercial software package
- Custom tools have been developed for data preparation, link storage and extraction
- Links for records belonging to one person are stored as a chain of pointers to records that can easily be added to or corrected
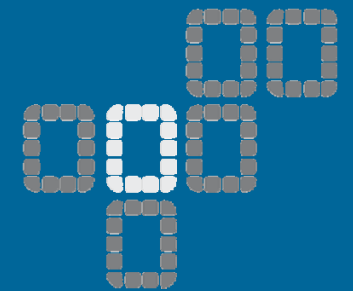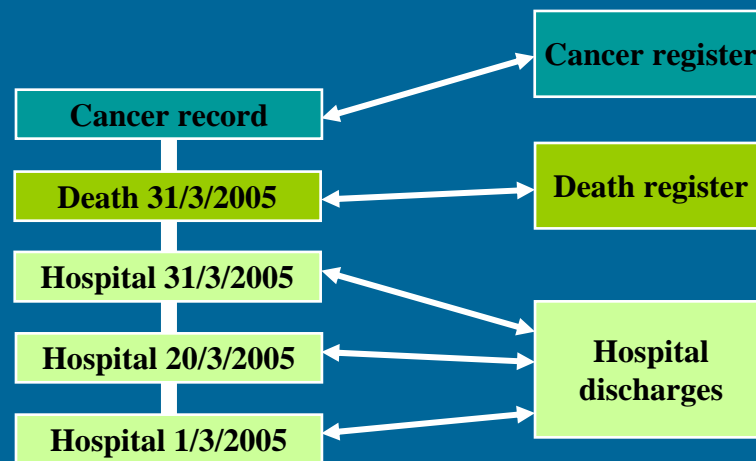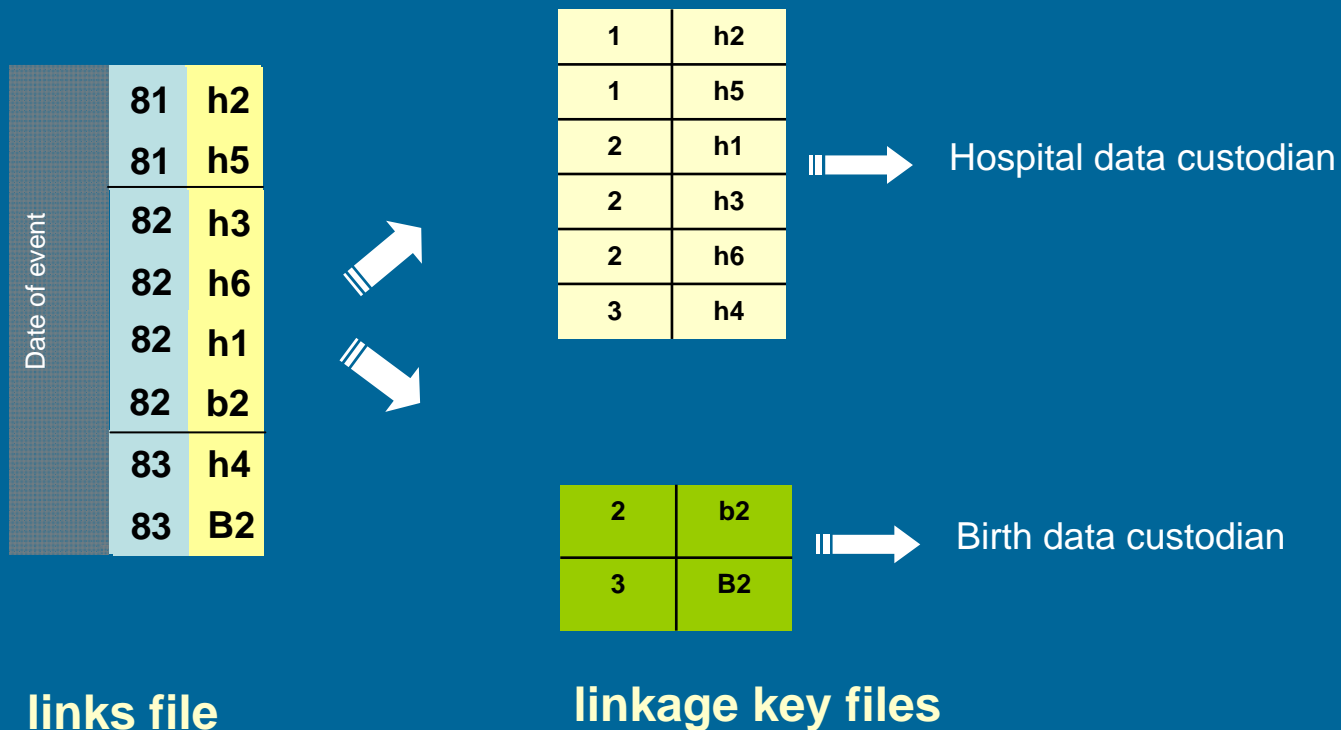
Links storage

Links found are loaded into the links file and are stored as a "chain of links". It is a database of links not a database of linked records

links file

unchanged demographics
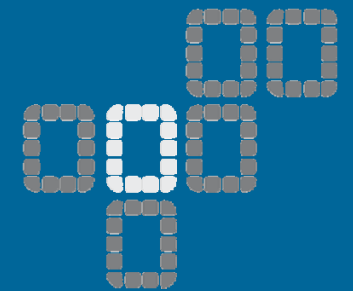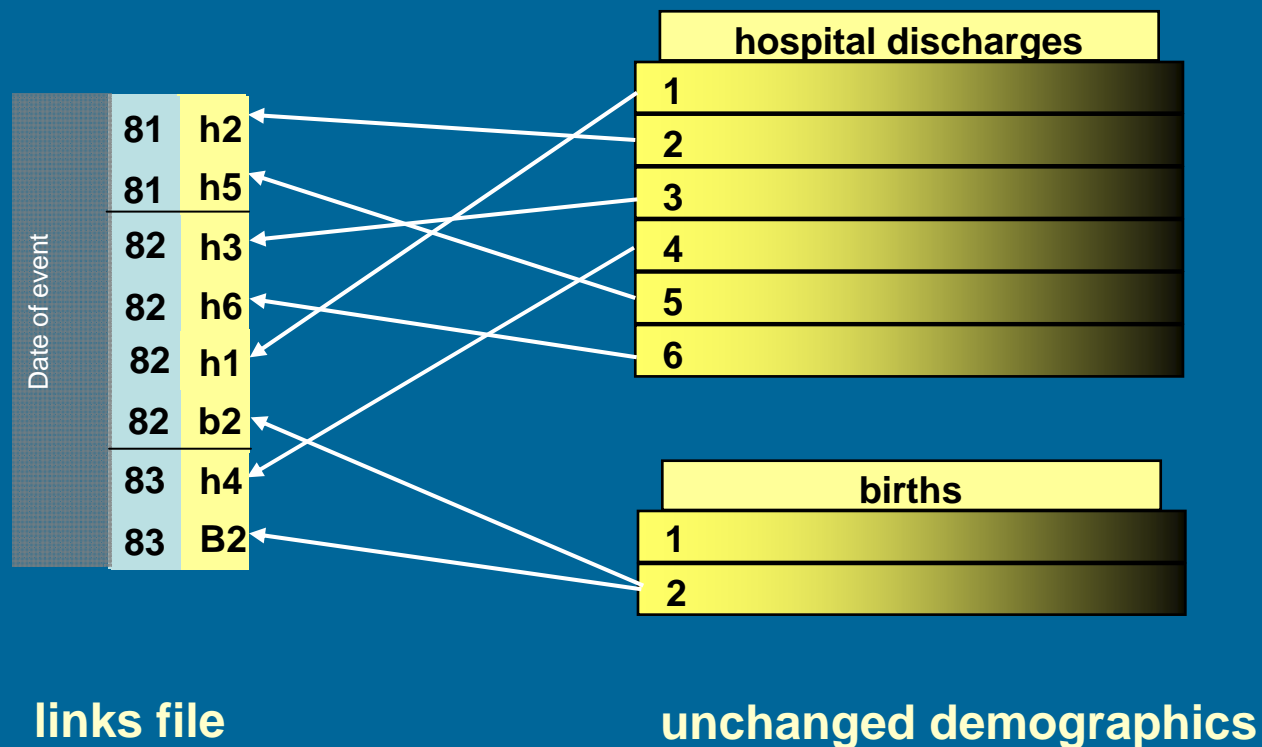
Linkage keys

Lists of source data record ids with a person id are produced from the WADLS links file and passed to data custodians for service data extraction

1	h2
1	h5
2	h1
2	h3
2	h6
3	h4

→ Hospital data custodian

Date of event

81	h2
81	h5
82	h3
82	h6
82	h1
82	b2
83	h4
83	B2

2	b2
3	B2

→ Birth data custodian

links file

linkage key files

Data linkage process - current

Morbidity → **Clean, standardise and load** →

Oracle database
- Morb tables
- Death tables
- Cancer tables

Death →

Cancer →

Extract demographic data →
- Morb
- Death
- Cancer

Linkage (Automatch)
- Morb <-> Morb
- Morb <-> Death
- Morb <-> Cancer
- Death <-> Cancer

Clerical review
- Links
- Links
- Links
- Links

Load links →

Oracle database
- *Links tables*

Raw data

Manual review

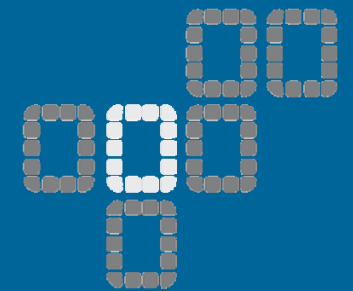Extract text-based demographic files for linkage

Set up and run linkage process
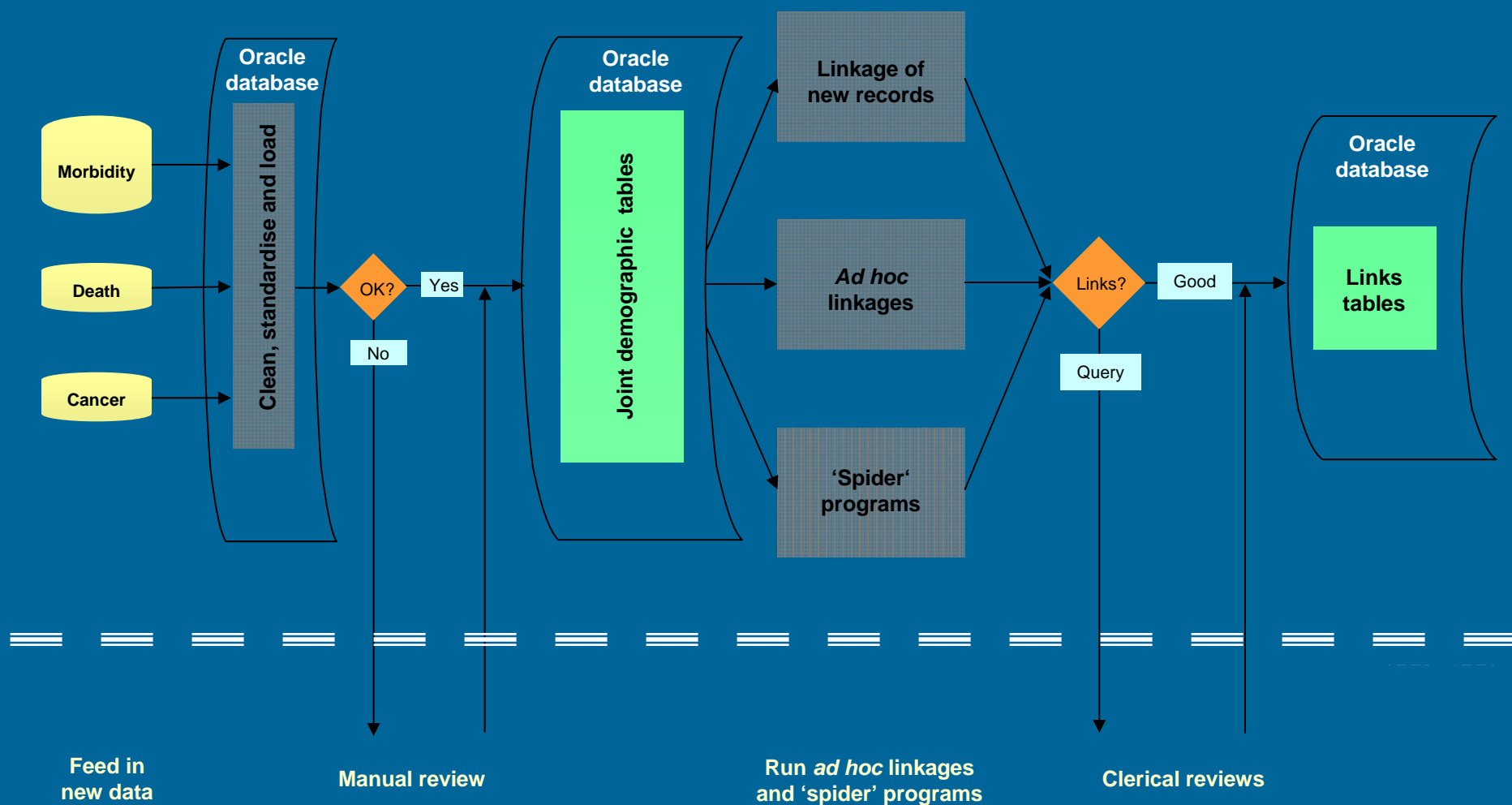
Clerical reviews

Extract text-based linkage results & load into links storage system

Data linkage process - current

- Multi-step,iterative process requiring constant operator intervention

- Command line driven with text based interfaces

- Every record in a batch must complete a phase before any records move to next step

- Approx 90% of records do NOT require any clerical cleaning or linkage

DLU linkage process - future

Morbidity

Death

Cancer

Oracle database

Clean, standardise and load

OK?

No

Yes

Oracle database

Joint demographic tables

Linkage of new records

Ad hoc linkages

'Spider' programs

Links?

Good

Query

Oracle database

Links tables

Feed in new data

Manual review

Run *ad hoc* linkages and 'spider' programs

Clerical reviews

DLU linkage process - future

- 3 main stages
  - Cleaning and loading of demographics
  - Linking
  - Storage of links
- Individual records will continue to next stage automatically unless clerical intervention is needed
- User interfaces GUI
- Linkage engine is being developed but could use other package
- Storage of links based on existing "chain of links" model currently in use
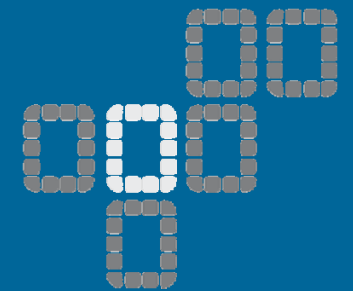
Data Linkage Branch
Public Health Division
Department of Health - Western Australia
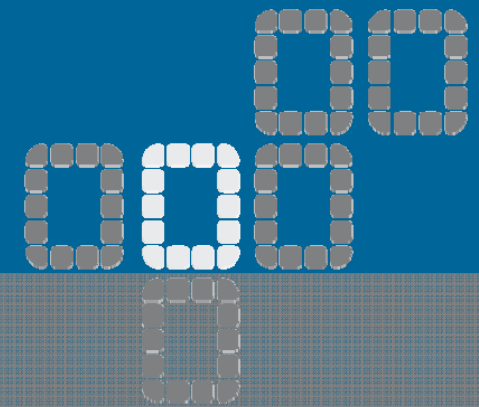
www.datalinkage-wa.org.au

Carol Garfield
Manager, Data Linkage Systems
Carol.Garfield@health.wa.gov.au
phone 08 9222 2371