



# Data Linkage

What it is and how we do it

**Raymond Daniel**

Manager, Data Linkage

Statistical Analysis & Linkage Unit

Statistical Services Branch, Queensland Health

# Overview



1

What is data linkage?

2

What is the MLF?

3

How is the MLF created?

4

What data are included in the MLF?



Please submit questions and comments via the chat

# What is data linkage?

Unique person identifiers are not often available within or across different data holdings in Queensland and Australia.

Data linkage is used to identify persons within or across different data holdings.

Uses person-level identifying information (such as name, date of birth, address) to determine which records pertain to a particular person.



# Our linkage teams

## Client Services

Supporting clients to access the data they need for research, insights and decision-making.

## Production Linkage

Creation of a Master Linkage File (MLF) and other linked data assets with enduring linkages between core health and other data collections.

## Request Linkage

Provision of data from the MLF and custom linkage for one-off research and departmental and other government agency requests.

# What is the MLF?

An **enduring map** of linked data enabling integration of unique event-level records at the person-level.

Maps a “**person**” to their records within and across data sources

**Linkage Key** – a numerical, enduring person-level identifier. A record’s linkage key changes only if its ownership changes.

**Source** – indicates the data source/ collection/ holding.

**Source Keys** – enable connections between the MLF and unique records in source data.

Linkage Key	Source	Source Keys
1000123456	QHAPDC	Pointer to a unique SC record
1000345678	QAS	Pointer to a unique QAS record
1000345678	EDC	Pointer to a unique EDC record
1000345678	QHAPDC	Pointer to a unique QHAPDC record
1000345678	SODC	Pointer to a unique SODC record
1000987654	RBDM-B	Pointer to a unique Birth Reg <sup>n</sup> record

# Mapping the Patient Journey

**MLF** – maps a “person” to their records

**Content data** – event-level information representing a person’s engagement with the system

Linkage Key	Source	Source Keys
1000123456	QHAPDC	Primary Key
1000345678	QAS	Derived Key
1000345678	EDC	PK
1000345678	QHAPDC	PK
1000345678	SODC	Derived Key
1000987654	RBDM-B	PK

Connects to

Queensland Ambulance Service record

Emergency DC record

Queensland Hospital Admitted Patient DC record

Specialist Outpatient DC record

Describes

Emergency Department Presentation

Patient Discharge



**1** Queensland Ambulance Service transport

**2** Emergency Department Presentation

**3** Hospital Admission (including across episodes and transfer between hospitals)

**4** Patient Discharge

**5** Outpatient Department Waiting List

# How is the MLF created?

- Most data sources are linked and **refreshed monthly**
- Extensive data preparation
- **Probabilistic linkage** using custom models
- **Human review** of uncertain matches (linkage “grey area”)
- Targeted **quality control** performed for each linkage cycle
- **Moderately conservative** linkage
  - Minimises false links while accepting a minor compromise in false non-links and keeping human review sustainable
  - Designed to support reliable **statistics, analytics, ML and AI**
  - **Not intended for direct clinical care**

# Probabilistic linkage custom models

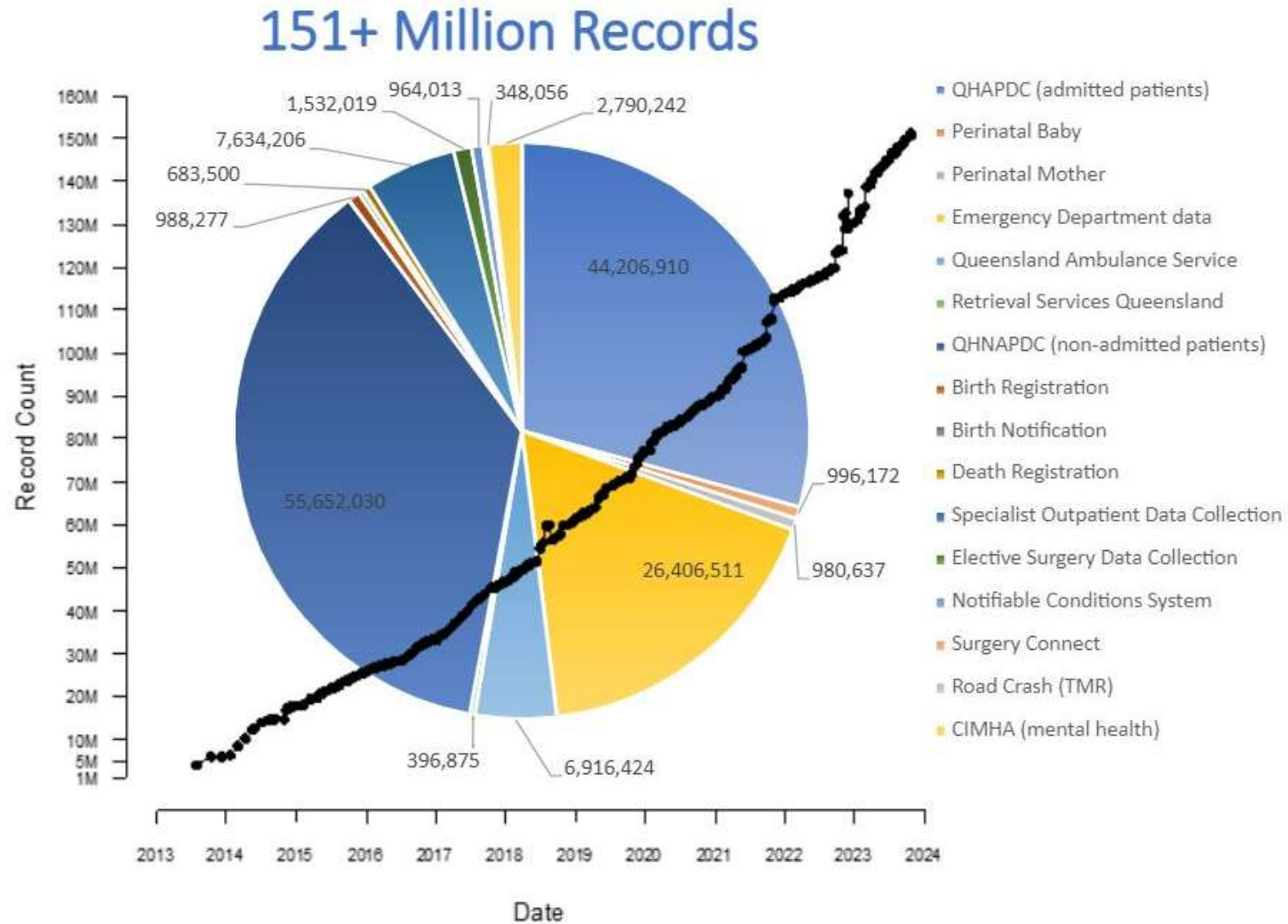
- Replicate **human decision-making**
- **Resilient** to some missing and inaccurate information
- Advanced blocking algorithm identifies potential matches to avoid the  $(n^2-n)/2$  problem
- Models use **~200 weighted Boolean tests** to compare person-level identifying information and calculate a probability of records matching
- Employ **exact and approximate string-matching**
- Trained using **supervised machine learning**
- Finely-tuned probability thresholds determine whether records match
- Targeted rules over-ride probability-based decisions
- Biconnected transitive closure determines groups of records (potentially) belonging to a person



# What data are included in the MLF?

	Data Source	Data From	Data To
ADMIT	Queensland Hospital Admitted Patient Data Collection (Public facilities)	01-01-2000	30-09-2023
ADMIT	Queensland Hospital Admitted Patient Data Collection (Private facilities)	01-07-2007	30-09-2023
QHNAP	Queensland Health Non-Admitted Patient Data	01-07-2016	03-10-2023
RBDM-BR	Birth Registration (Registry of Births, Deaths and Marriages)	01-07-2007	27-09-2023
	Birth Notification (Registry of Births, Deaths and Marriages)	01-01-2017	29-09-2023
RBDM-D	Death Registration (Registry of Births, Deaths and Marriages)	01-01-2000	30-09-2023
QAS	Queensland Ambulance Service	01-12-2015	31-07-2023
RSQ	Retrieval Services Queensland (Includes Brolga, and the historical Queensland Neonatal Emergency Transport Service and Clinical Coordination Retrieval Information System holdings)	16-10-2017 (Brolga)	11-09-2023
CIMHA	Consumer Integrated Mental Health Application	01-01-2010	24-09-2023
EDC	Emergency Data Collection	01-07-2015	31-08-2023
QHED	Public Emergency Department holding	01-07-2008	30-06-2015
SODC	Specialist Outpatient Data Collection	01-06-2015	06-10-2023
ESDC	Elective Surgery Data Collection	01-07-2015	31-08-2023
NOCS	Notifiable Conditions System	01-01-2010	30-05-2021
PDC	Perinatal Data Collection (Public and Private facilities)	01-07-2007	30-09-2023
SC	Surgery Connect	(Undefined)	05-10-2023
TMR	Department of Transport and Main Roads (Road Crash Data)	01-01-2010	30-06-2022

# What data are included in the MLF?



# What data will soon be included in the MLF?

Data sources to be added in 2024:

- **Queensland Master Patient Index** – to support analyses across clinical and administrative datasets
- **Notifiable Conditions System** – resume linkage to support ongoing health surveillance and analysis
- **Medicare Consumer Directory** – to support efficient use of existing linked data for cross-jurisdictional projects and national analytic assets



# Key takeaways about the MLF

- Linkage map with **enduring Linkage Keys**
- Coverage is **broader than QH Public** facilities and systems
- Coverage is continually **expanding**
- Quality linkage developed for **statistics, insights, etc.**



Thank  
you



[dlq@health.qld.gov.au](mailto:dlq@health.qld.gov.au)