Queensland Health

Introduction to the Master Linkage File

Raymond Daniel
Statistical Analysis and Linkage Unit
Statistical Services Branch

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

Queensland Government

Presentation outline

- Overview the Master Linkage File (MLF)
- Data source integration
- Routine data linkage cycle
- Yearly quality assessment
- Future developments

What is the MLF?

- A map of an individual's journey through the Queensland health system and related services
- Maintained by routinely bringing together information from multiple data collections and registries
- A simple "mapping" table
- Contains enduring linkages between an individual's event-level records

Maps records to “persons”

Distinct key = distinct “person”

Identifies a record in source data. Facilitates joins to the source.

Linkage Key	Source	Source Key/s
12345678AAA	ADMIT	References an Admitted Patient record
12345678AAA	ADMIT	References an Admitted Patient record
98765432BBB	PERIB	References a Perinatal Baby record
98765432BBB	ADMIT	References an Admitted Patient record
98765432BBB	BIRTH	References a Birth Registration record
34567890GGG	QAS	References a QLD Ambulance record
34567890GGG	EDC	References an Emergency Dept. record

What the MLF is not

The Master Linkage File is not

- A repository of identifiable data.
  - Identifiable data are not available from the mapping table/s
- A repository of content data.
  - Content data are available only from the source data collections

Data Collections - linked

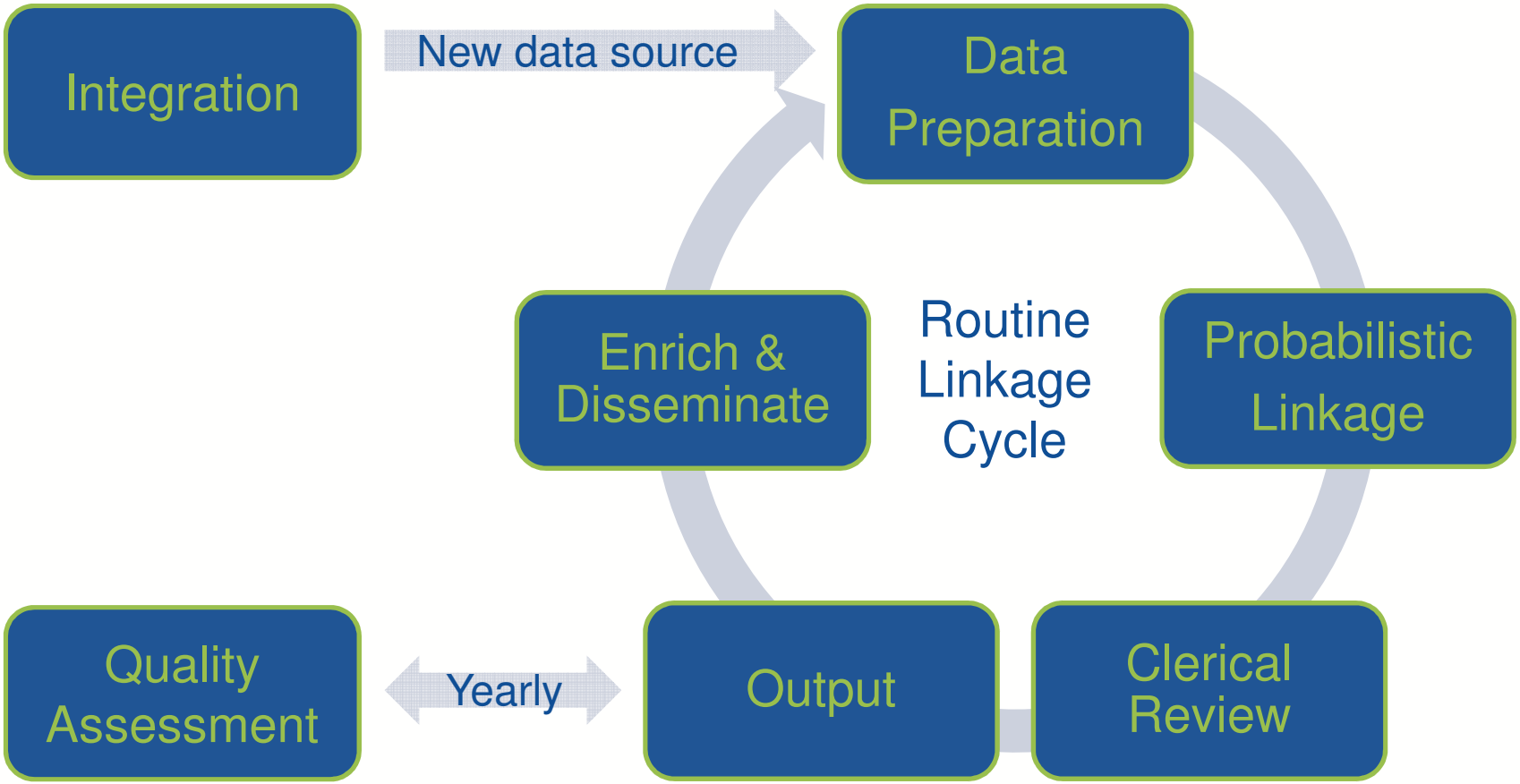
Data Collections	Coverage
Queensland Hospital Admitted Patient	Public: Jan-2004 to current Private: Jul-2007 to current
Emergency Department	Jan-2012 to current
Perinatal (Mothers & Babies)	Jul-2007 to current
Birth Registration	Jul-2007 to current
Death Registration	Jan-2004 to current
Elective Surgery Waiting List	Jul-2014 to current
Outpatient Waiting List	Jul-2015 to current

The MLF contains approx. 32 million records, with current growth at approx. 800,000 records per month.

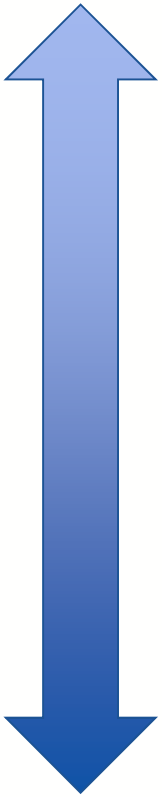
Data collections - pending

Data Collections	Status
QLD Ambulance Service	Integration commenced
Surgery Connect	Integration commenced
Notifiable Conditions System	Approved
Vaccination Information	Approved
Non-Admitted Patient (Specialist Outpatient)	Approved
Air Retrievals	Approved
Education (AEDC, NAPLAN)	Approved/Negotiating access

High-level routine linkage process



Integration



Stakeholder engagement

- Custodians: Data content, quality, timeliness, dynamic?..
- Users: Linkage quality

Data exchange

- Implement systems for data receipt and dissemination

Data quality assessment

- Data profiling and issue management

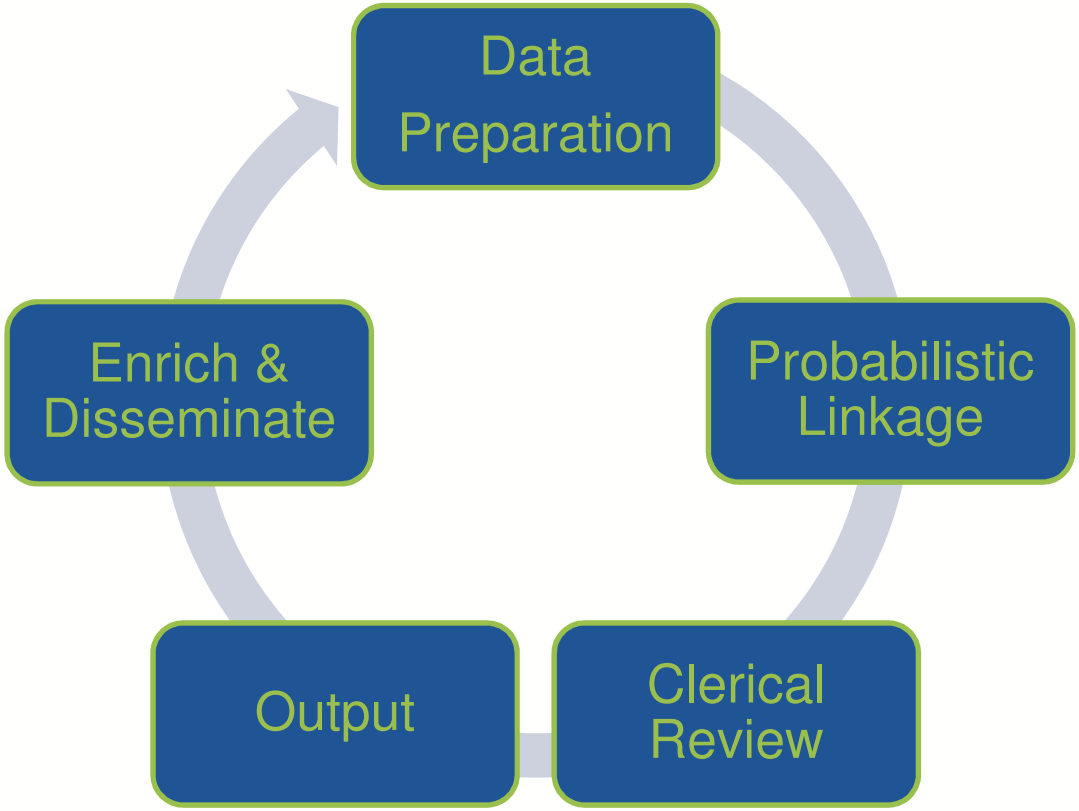
Database & script development

- Data preparation processes and linkage infrastructure

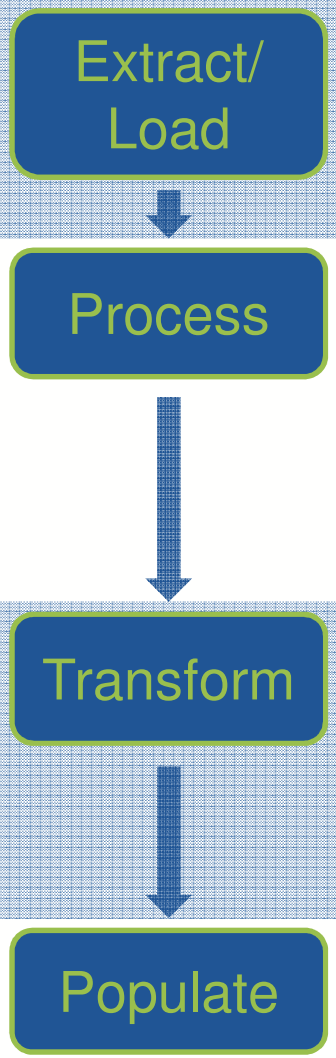
Linkage model development

- Ensure linkage quality meets stakeholder expectations

Regular linkage cycle



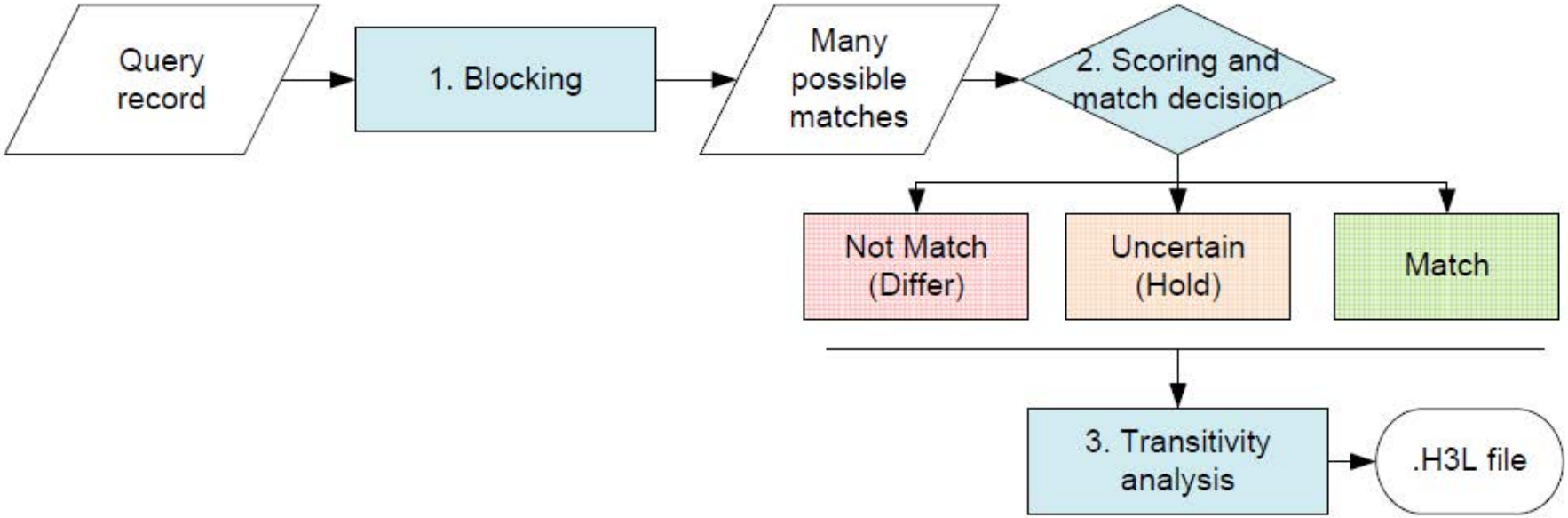
Data preparation



- 2+ extractions per month
- Extract after data due at SSB
- Process dynamic data
  - Allows use of current “in-work” data
  - Ensures MLF consistent with source
  - Meets emerging needs
- Standardise phone numbers
- Comprehensively parse names
- Map to QHDD formats and values
- Make data available to engine

Probabilistic linkage

- Minimise human review while maintaining quality
- 3-Stage process performed by the ChoiceMaker linkage engine



Probabilistic linkage

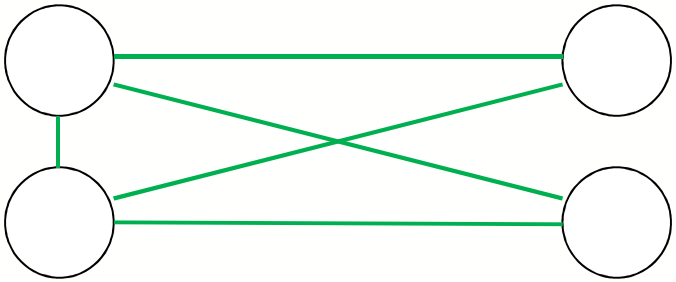
1. **Blocking**
   - Creates “blocks” of records that possibly match to each other to minimise unnecessary comparisons
   - ChoiceMaker algorithm and 17 blocking fields to efficiently minimise false negative errors
2. **Scoring and match decision**
   - Custom model determines whether records match, differ or undecided (hold)
   - 150 weighted tests used to calculate probability
   - Tuned thresholds assign the decision
   - 14 deterministic rules can over-ride decisions

Probabilistic linkage

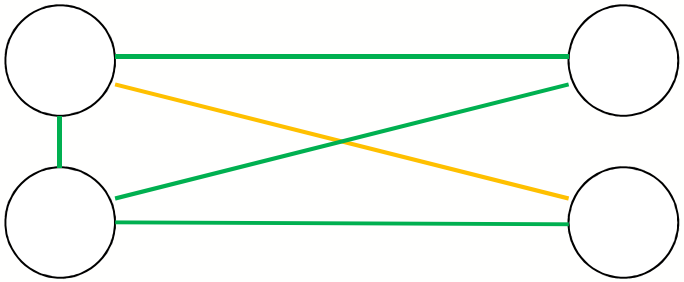
3. Transitivity analysis

- Creates two types of linkage groups: merge and hold
- Bi-connected transitive closure of records creates merge groups: records will be assigned same key
- Incomplete transitive closure of related records creates a hold group for clerical review (grey area)

Merge group



Hold group



Regular linkage cycle

Clerical review of uncertain links

Human review of hold groups (grey area)

- In-house developed tool for review
- 1,000 to 20,000 hold groups per linkage
- Requires sustained focus and quick decision-making
- Some decisions are not clear-cut
- Interrogation of source data often required
- Maintaining consistency is paramount
- A tough job, but we have a very capable team!

Regular linkage cycle

Output the linkage results

- Results written to MLF draft tables
- Routine quality checks performed
- Errors corrected
- Archived for reference

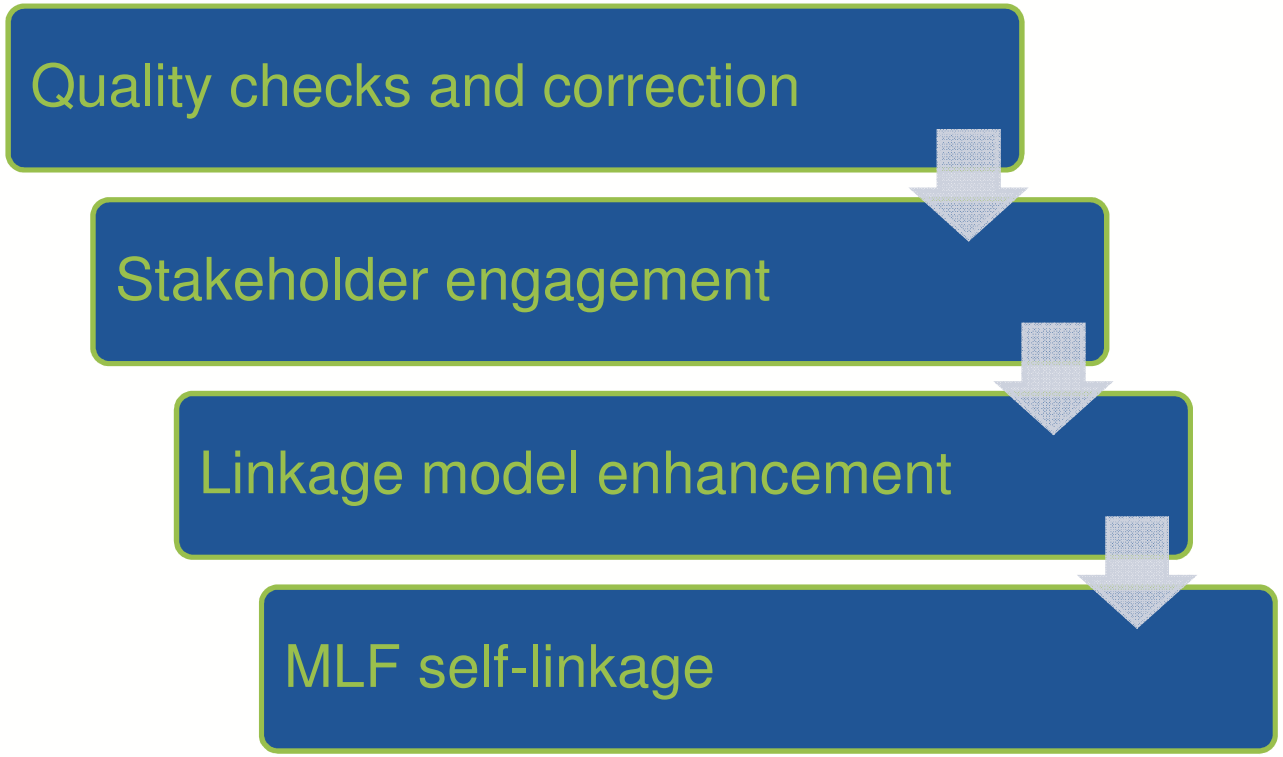
Regular linkage cycle

Enrich and disseminate linked data

- Refresh data for QH users – direct access to MLF
- Routine data enrichment and dissemination
  - Collection reconciliation (PDC:QHAPDC)
  - Secondary linkages & data provision (QCOR)
  - Linked files (CALF, deceased list…)

Yearly quality assessment

Quantify and improve MLF quality



Yearly quality assessment

Quality checks and correction

- Quantify and correct subsets of linkages known or suspected to be problematic
  - Approx. 25 subsets checked
  - E.g. Twins falsely linked
- Accuracy
  - 10/1000 errors per subset (ideally 5/1000)
  - 5/1000 errors per MLF

Yearly quality assessment

Stakeholder engagement

- Gain consensus regarding linkage quality
- Often relates to linkage error bias: “how conservative?”

Enhance the linkage model

- Modelling to remove root-causes of significant errors and implement stakeholder feedback
- Improves the quality of subsequent linkages

Yearly quality assessment

Self-link the MLF

- Applies model enhancements to previously linked data
- Corrects false-negative errors by merging disparate linkage groups

Future developments

- Continued integration of new data sources
  - Queensland Ambulance Service
  - Surgery Connect
- Implement enduring linkage keys
- Operational continuity and efficiency
  - Migrate to the current version of the ChoiceMaker linkage engine
  - Infrastructure enhancement
- Continued quality and process improvement – the never ending quest

Contacts and information

https://www.health.qld.gov.au/hsu/link/datalink.asp