LOCAL LEVEL DATA LINKAGE

Probabilistic Record Linkage Algorithm: a pilot study

Data linkage plays a key role in improving primary health outcomes by enabling individual health trajectories.

Deterministic record linkage is often not possible in the absence of a common identifier between multiple data sources. This study used a local-level probabilistic approach to link records in different settings.

A complete picture of individual trajectories help clinicians and policy makers to reduce adverse health outcomes and improve health and wellbeing in the community

KEY FINDINGS AND FUTURE WORK

- Western Sydney Local Health District data sample included 2681 records, and 796 records matched with Western Sydney Primary Health Network.
- Linkage accuracy exceeded 95% in terms of sensitivity and specificity.
- Future research is underway to scale this work to a wider cohort.

METHODOLOGY

- A hashed key will be generated based on the identifiable information before data being shared.

First name	Last name	Date of birth	Sex
Sithum	Munasinghe	01jan1900	M
Shahana	Ferdousi	01jan1901	F
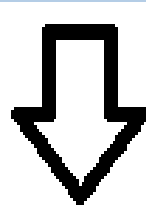
| ID	hash
1	3542*562%2544527#8426236$72%36*1623$623$73&8!9825%9234*$924537%26$9823&973*73
2	13%25437&24736*147$98235&8!25437$924537%26$734%2352$2753782!57384*

- Probabilistic linkage algorithms will be run for hashed data.
- Linkage quality checks will be conducted to identify optimal linkage threshold (this will only be conducted for the first time).
- A linkage between PHN key and other organisation unique keys will be established.
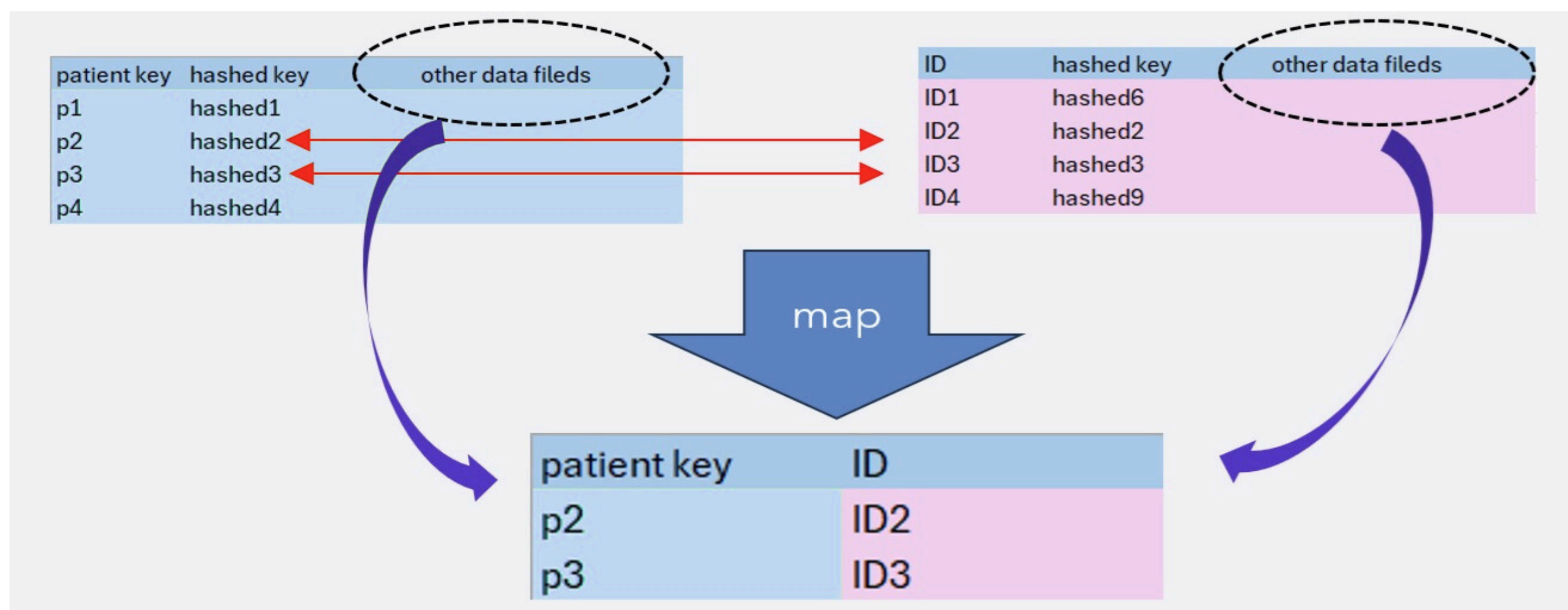
RESULTS

Probability threshold	linked pairs	True match	Non-match
100%	177 (22.2%)	177 (81.6%)	0 (0%)
95-99.99%	35 (4.4%)	35 (4.4%)	0 (0%)
90-95%	6 (0.8%)	4 (0.8%)	2 (0.3%)
85-90%	21 (2.6%)	1 (0.5%)	20 (3.5%)
80-85%	152 (19.1%)	0 (0%)	152 (26.3%)
75-80%	299 (37.6%)	0 (0%)	299 (51.6%)
below 75%	106 (13.3%)	0 (0%)	106 (18.3%)
Total	796	217	579

Note: 20 records per each probability cut-oof were selected for clerical review
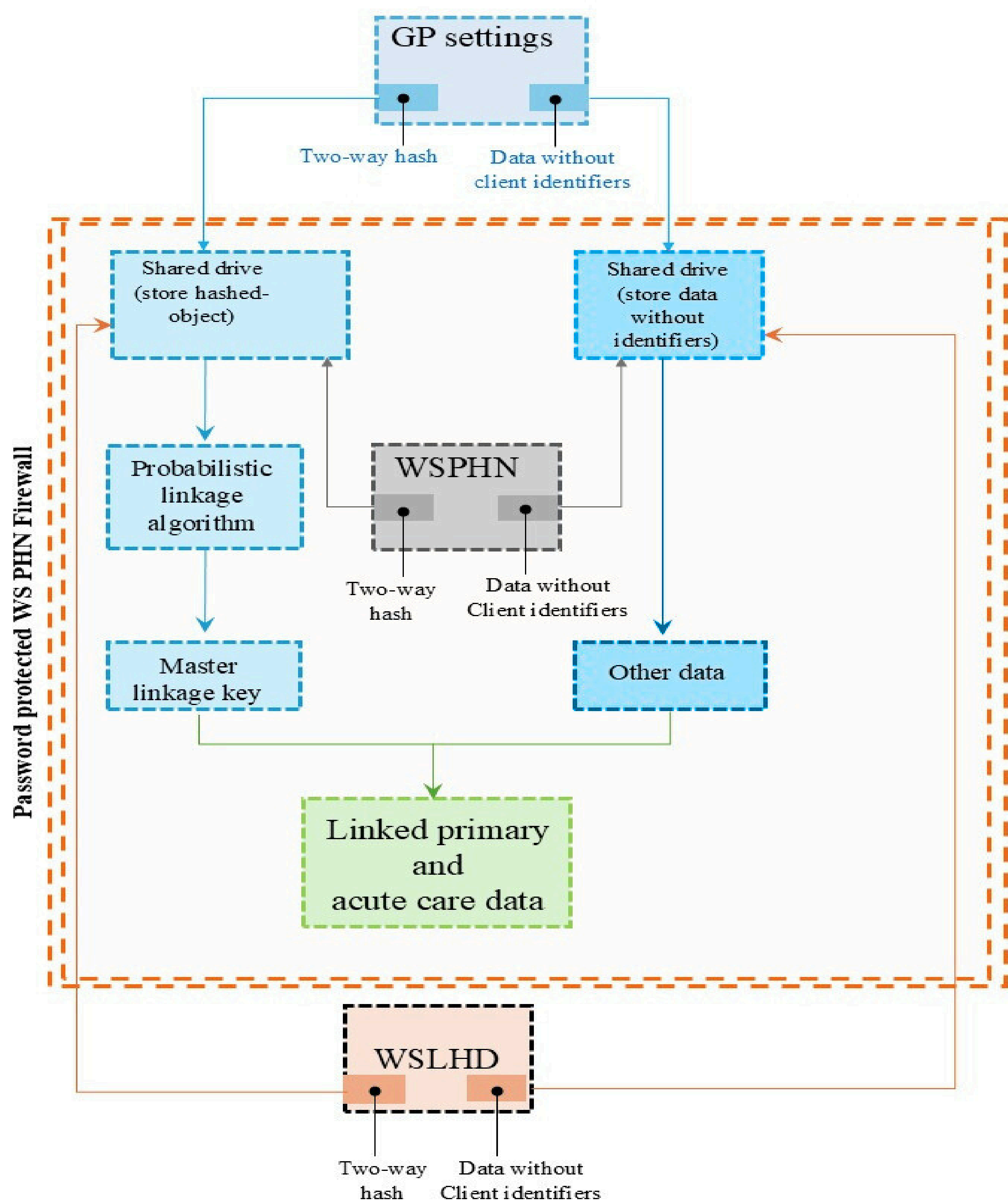
Authors:

Dr Sithum Munasinghe, Dr Shahana Ferdousi, Maxwell Osaghae, David Pryce, Piying Shi, Prof. Vlasios Brakoulias, and Prof. Andrew Page

e

References
Sariyar M, Borg A. Package 'RecordLinkage'. 2020. Contiero P, Tittarelli A, Tagliabue G, et al. The EpiLink record linkage software. Methods of Information in Medicine. 2005;44(01):66-71. Winkler WE. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. 1990

Karr AF, Taylor MT, West SL, et al. Comparing record linkage software programs and algorithms using real-world

data. PloS One. 2019;14(9):e0221459. Morgan AS, Marlow N, Costeloe K, Draper ES. Investigating increased admissions to neonatal intensive care in England between 1995 and 2006: data linkage study using Hospital Episode Statistics. BMC Medical Research Methodology. 2016;16(1):1-11.