

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Implementing privacy-preserving record linkage: welcome to the real world

Boyd, James^{1*}, Ferrante, Anna¹, Brown, Adrian¹, Randall, Sean¹, and Semmens, James¹

¹Curtin University

Objective

While record linkage has become a strategic research priority within Australia and internationally, legal and administrative issues prevent data linkage in some situations due to privacy concerns. Even current best practices in record linkage carry some privacy risk as they require the release of personally identifying information to trusted third parties. Application of record linkage systems that do not require the release of personal information can overcome legal and privacy issues surrounding data integration. Current conceptual and experimental privacy-preserving record linkage (PPRL) models show promise in addressing data integration challenges but do not yet address all of the requirements for real-world operations. This paper aims to identify and address some of the challenges of operationalising PPRL frameworks.

Approach

Traditional linkage processes involve comparing personally identifying information (name, address, date of birth) on pairs of records to determine whether the records belong to the same person. Designing appropriate linkage strategies is an important part of the process. These are typically based on the analysis of data attributes (metadata) such as data completeness, consistency, constancy and field discriminating power. Under a PPRL model, however, these factors cannot be discerned from the encrypted data, so an alternative approach is required. This paper explores methods for data profiling, blocking, weight/threshold estimation and error detection within a PPRL framework.

Results

Probabilistic record linkage typically involves the estimation of weights and thresholds to optimise the linkage and ensure highly accurate results. The paper outlines the metadata requirements and automated methods necessary to collect data without com-

promising privacy. We present work undertaken to develop parameter estimation methods which can help optimise a linkage strategy without the release of personally identifiable information. These are required in all parts of the privacy preserving record linkage process (pre-processing, standardising activities, linkage, grouping and extracting).

Conclusion

PPRL techniques that operate on encrypted data have the potential for large-scale record linkage, performing both accurately and efficiently under experimental conditions. Our research has advanced the current state of PPRL with a framework for secure record linkage that can be implemented to improve and expand linkage service delivery while protecting an individual's privacy. However, more research is required to supplement this technique with additional elements to ensure the end-to-end method is practical and can be incorporated into real-world models.

*Corresponding Author:

Email Address: j.boyd@curtin.edu.au (J. Boyd)

