

## Privacy-Preserving Record Linkage: An international collaboration between Canada, Australia and Wales

Pow, Conrad<sup>1\*</sup>, Iron, Karey<sup>1</sup>, Boyd, James<sup>2</sup>, Brown, Adrian<sup>2</sup>, Thompson, Simon<sup>3</sup>, Chong, Nelson<sup>1</sup>, and Ma, Charlotte<sup>1</sup>

<sup>1</sup>Institute for Clinical Evaluative Sciences

<sup>2</sup>Curtin University

<sup>3</sup>Swansea University

### Objectives

Linkage of “big data” can provide the answers to a variety of health questions that benefit the delivery of patient care, impact of policies, system planning and evaluation. In some jurisdictions, legal and operational barriers may prevent data linkage for research and system evaluation. Collaboration between international research institutions in Canada, Australia and Wales was formed at the Farr Institute International Conference in 2015. This partnership will test privacy-preserving record linkage (PPRL) techniques for linkage accuracy on real datasets held in a Canadian data repository.

### Approach

Bloom filter PPRL techniques have been incorporated into a prototype linkage system. Evaluations on probabilistic linkage using Bloom filters method have shown potential for large-scale record linkage, performing both accurately and efficiently under experimental conditions. The prototype will be used to evaluate the Bloom filter PPRL techniques in 3 phases. Phase 1: 3 tests using simulated data relating to 20 million individuals will be matched to a sub-cohort of 1 million individuals. Phase 2: 100,000 people from hospital inpatient records will be matched to 18 million people in a health system registration file.

These tests will inform whether the method can achieve high levels of privacy protection without negatively impacting performance and linkage quality. Performance indicators include match rate and processing efficiency based on record volumes.

### Results

Linkage quality will be assessed by the number of true matches and non matches identified as links and non-links. This method

will be evaluated using synthetic and real-world datasets, where the true match status is known. Initial performance testing linked a file of 3,000 records to 30,000 with a 100% match result. Subsequent test phases as above will continue to be evaluated and these results will be presented.

### Conclusion

Completion of the phased tests will confirm the ability to link datasets while preserving privacy. This international collaboration will expand the utility of this prototype linkage system and expand the global knowledge bank focusing on PPRL methods in general. It will also inform how to adapt to local requirements by providing a solution to many common legal and administrative challenges.

\*Corresponding Author:

Email Address: [conrad.pow@ices.on.ca](mailto:conrad.pow@ices.on.ca) (C. Pow)

