

# International Journal of Population Data Science

Journal Website: [www.ijpds.org](http://www.ijpds.org)



## An effective privacy enhanced interface to support record linkage decisions

Kum, H-C<sup>1</sup>, Ilangoan, G<sup>1</sup>, Li, Q<sup>1</sup>, Li, Y<sup>1</sup>, and Ragan, E<sup>1</sup>

<sup>1</sup>Texas A&M University

### Introduction

Privacy enhanced technologies (PET) are those that measure and protect privacy by preventing unnecessary use of personal data without loss of the functionality of the information system. In practice, implementing such a system requires fine-grained access control so that access can be granted in smaller chunks of data.

exposure of PII to people when resolving ambiguous linkages without compromising linkage quality. This research points to a new direction for PET in record linkage beyond encryption.

### Objectives and Approach

In record linkage, PET to date has mostly meant separation of identifiers and sensitive information to allow access to only the necessary part. Moving beyond this current norm, we have designed a privacy enhanced interface to support linkage that discloses only the needed information at the sub variable level, when needed to make good decisions, to reduce exposure of personally identifiable information (PII). The system allows for access to PII both at (1) cell level (e.g., only names of needed people are released) or (2) sub-cell level (e.g., only part of a name, suffix or characters, is released).

### Results

In a user study (N=104) where participants tried to link complicated situations (e.g. twins, Sr/Jr, change of last name) using the interface, we found that users given fully masked data, 0% of information disclosed, were still able to get 75% accuracy using supplemental visual markup. The markups depict data discrepancies such as swapped first and last names, transposed characters, different characters and missing data. More importantly, with this effective interface, we found that there were no statistical difference in accuracy of linkage (84%) or time taken between users with access to all data and those with access to only 30% of the data. We have released a tutorial where users can experience balancing between information disclosure and accuracy of results on sample data.

### Conclusion/Implications

Privacy is a major public concern when PII is legitimately accessed to link data. Our study demonstrates that a well-designed privacy enhanced interface can significantly reduce

