

# International Journal of Population Data Science

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



Swansea University  
Prifysgol Abertawe

## Preserving Privacy in the Face of Clients with Different Data Needs Using the NGLMS

Radbone, Chris<sup>1\*</sup> and Farrow, James<sup>1,2</sup>

<sup>1</sup>SA NT DataLink

<sup>2</sup>Farrow Norris

### Objectives

SA NT DataLink's Next Generation Linkage Management System (NGLMS) provides a novel approach to handling privileged or sensitive data for certain projects without having to replicate or duplicate databases and work to protect privacy. The NGLMS is a collection of records (nodes) and relationship (edges) that forms a graph (in the computer science sense) and is designed to support a mix-and-match and layered approach to data linkage projects. The NGLMS allows for the needs of different clients to be managed from the one graph data set while preserving privacy and honouring the requirement to protect sensitive information without having to relink or duplicate data.

### Approach

The NGLMS uses a layer-based approach to project description and design. Projects, a specific data linkage request for example, are composed of various data layers. The data layers consist of data sets, link information in the form of pairwise relationships. These layers are coupled with quality information, e.g. acceptable similarity thresholds and/or the types of relationships to consider as 'linking' two records, to construct an effective virtual data set which may be different for each project. A project can be constructed by composing existing linkage data (where it already exists) without having to perform new linkage comparisons.

### Results

A case study will be discussed where a data set containing extremely sensitive information (record pairings revealing name changes due to family court proceeding and protection orders) was received for incorporation into the data pool. This information is sensitive for which the particular data custodian who

supplied the information would wish to have honoured by only incorporating their records for approved analysis, and otherwise excluded for other non-authorised analysis. By placing these data into a separate layer to be included in some projects and not others the sensitive nature of the data can be accommodated and its effects 'turned on and off' at will.

### Conclusion

The flexible on-demand nature of data extraction and late clustering in the NGLMS Graph based approach the linkage allows for ad-hoc project construction and the dynamic inclusion and exclusion of data without the overhead of relinking data.

\*Corresponding Author:

Email Address: [chris.radbone@unisa.edu.au](mailto:chris.radbone@unisa.edu.au) (C. Radbone)

