

International Journal of Population Data Science

Journal Website: www.ijpds.org



“Hiding in the Crowd”: defining conceptual models for privacy-preserving linkage of place-based and personal data for public benefit research.

Andy Boyd¹, Chris Dibben², Rich Fry³, Pia Hardelid⁴, Cassie Smith⁵, Edel McNamara⁵, Bethan Davies⁶, Daniela Fecht⁶, Ruth Blackburn⁴, and Ruth Gilbert⁴

¹University of Bristol

²University of Edinburgh

³Swansea University

⁴UCL

⁵Health Data Research UK

⁶Imperial College London

Objective

To address the challenge of linking identifiable but low-sensitivity place-based data with high-sensitivity, de-identified, health and socio-economic records. While also minimising the financial and carbon costs of computationally intensive environmental modelling.

Approach

For domain and legal experts, TRE leads and public contributors to co-develop a trusted system-wide governance model for linking these data.

Results

We identified the UK's 'Unique Property Reference Numbers' (UPRNs) - a geo-coordinated national property ID number - as ideally suited for linking individuals to households and households to place-based data. Under Data Protection laws, UPRNs are considered inherently identifiable, yet public-domain research use is permitted as the data have low sensitivity and there is broad societal acceptance. Once linked with individuals' records, additional controls are needed to preserve privacy. We developed a workflow for modelling place-based data at a national level by a single agency and then, using trusted third parties, de-identifying and sharing minimised derived extracts into TREs for linkage and analysis. To maintain confidentiality, the UPRNs of the population of interest are masked by matched 'control' UPRNs drawn from the wider population. We will discuss permutations of the model, proof-of-concept implementations in UK TREs and public views on the models proposed.

Conclusions

The challenge can be addressed by minimising the data flowing between national generators of place-based data and TREs using trusted third parties and masking data to maintain confidentiality.

Implications

Relatively small adjustments to existing TRE workflows principles could enable new research possibilities without requiring re-evaluation of current governance norms or high-cost infrastructure changes.



Temporary page!

$\LaTeX$ was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because $\LaTeX$ now knows how many pages to expect for this document.