

International Journal of Population Data Science

Journal Website: www.ijpds.org



Linking vital events to census records for accurate and efficient large-scale population reconstruction

Sumayya Ziyad¹, Peter Christen², and Eilidh Garrett³

¹Australian National University, Canberra, Australia

²University of Edinburgh, Edinburgh, United Kingdom; Australian National University, Canberra, Australia

³University of Edinburgh, Edinburgh, United Kingdom

Certificates of birth, death, and marriage provide information about key vital events in people's lives. Census records, on the other hand, provide snapshots of individuals, families, and households at certain points in time. Registers of vital events and census records both cover large portions of a population, and in combination they can help to reconstruct populations over time. Such integrated views of population data can provide crucial insights into the processes underlying demographic trends. In this work we investigate how record linkage methods applied to vital events can be complemented with linkages to census records, and vice versa. We develop bespoke linkage methods starting with exact matching followed by approximate matching. For each method we evaluate linkage quality (the precision of the identified links) and linkage completeness

(the recall of the identified links). We report on two studies using historical Scottish data sets for which we have ground truth data available. Based on the results achieved we then explore how our methods can be applied on a large population database spanning over a century and containing more than 100 million person records. Initial experiments using exact matching on individual and parental names to link birth certificates to four snapshots of census records achieved an average precision of 0.93, demonstrating that the linkage between newborns and their census records can be highly accurate. The average recall of 0.57 for exact matching highlights the need to employ approximate matching methods to facilitate better linkage of those records that contain errors and discrepancies.

