

Knowledge production through record linkage: reconstructing sailors' careers for insight into migrants' opportunities in and contribution to the Dutch economy, c. 1700-1800

Rutger van Koert, Huygens ING

Jelle van Lottum, Huygens ING

Lodewijk Petram, Huygens ING

We apply innovative semi-automatic methods for record linkage on biographical data gathered from various large, heterogeneous historical data sets to gain historical insight into a topic that is currently hotly debated in both the political and public arena: what is the economic contribution of migrant workers on a recipient economy?

We present preliminary results from ongoing research into the eighteenth-century Dutch maritime sector. This was a key sector of the economy, characterised by a high level of migrant participation and well-known today for leaving rich sources. Our approach to uncover the link between migration and economic advancement centres on two aspects of what is generally referred to as job mobility: promotion and job switching. By identifying those individuals that during their careers experienced promotion and/or job switching we seek to lay bare the extent to which human capital levels of migrants and native workers differed, and what opportunities for social advancement were available in the eighteenth-century Dutch maritime labour market.

The data on migration and job mobility are obtained through record linkage of biographical data contained in muster rolls of the Dutch East India Company (almost 800,000 records), and transcripts of interrogations of crew members of Dutch merchant marine by the English admiralty (c. 15,500 records). We use various methods to measure name similarity (trigram matching, Levenshtein distance, (Double) Metaphone) and filter out irrelevant results by applying a set of rules based on dates (for example, a person cannot sail out before birth or after death). Finally, we use the sailors' places of birth as a check on matches, first by looking for exact matches between standardised toponyms, but also by geo-grouping locations and measuring spelling similarity of the original toponyms.

We reflect on our matching methodology and give a first impression of the record linkage tool that has been developed for our project and that is projected to be integrated into the CLARIAH digital infrastructure. Furthermore, we discuss the data model for the output of the linkage procedure, i.e. a large collection of mini-biographies of maritime workers, containing information on e.g. place of birth, places of residence, and job positions held at various points in time, from which we extract geographical mobility and career events. Finally, we show how we integrated the data into the CLARIAH infrastructure and subsequently performed our statistical analyses.

Our short paper fits seamlessly with the conference theme. The digital techniques and methodology applied in our project open up possibilities for employing vast historical data collections – known by historians for a long time, but difficult to handle because of their size – in large-scale analyses that may be instrumental in answering urgent, challenging research questions.