Population Reconstruction

Gerrit Bloothooft · Peter Christen
Kees Mandemakers · Marijn Schraagen
Editors

Population Reconstruction

Springer

Editors
Gerrit Bloothooft
Utrecht University
Utrecht
The Netherlands

Kees Mandemakers
International Institute of Social History
Amsterdam
The Netherlands

Peter Christen
The Australian National University
Canberra, ACT
Australia

Marijn Schraagen
Leiden University
Leiden
The Netherlands

ISBN 978-3-319-19883-5 ISBN 978-3-319-19884-2 (eBook)
DOI 10.1007/978-3-319-19884-2

Library of Congress Control Number: 2015942214

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

People shape societies. They are linked to each other by family ties and networks with social, economic and religious dimensions. People live together in households and form communities. Some own a house, land and other properties, often related to their profession. And all this is in continuous change. People are born, marry, have children and die, and they change houses and addresses, and build careers. For the study of a society in all aspects, people are at the heart of the problem and should be known in the context of their complex relationships. Even today, it is not easy to get this information in an all-enfolding way, but for populations in the past, it is a real challenge. And that is what this book is about. The book addresses the problems that are encountered, and solutions that have been proposed, when we aim to identify people and to reconstruct populations under conditions where information is scarce, ambiguous, fuzzy and sometimes erroneous.

It is not a single discipline that is involved in such an endeavour. Historians, social scientists, and linguists represent the humanities through their knowledge of the complexity of the past, the limitations of sources and the possible interpretations of information. The availability of big data from digitised archives and the need of complex analyses to identify individuals require the involvement of computer scientists. With contributions from all these fields, often in direct cooperation, this book is at the heart of digital humanities and hopefully a source of inspiration for future investigations.

The process from handwritten registers to a reconstructed digitised population has three major phases which shape the three sections of this book. The first phase is that of data transcription and digitisation while structuring the information in a meaningful and efficient way. Little of this phase can be automated. With archives that comprise easily tens of millions of records, the help of volunteers for transcription and digitisation is indispensable, but requires a rigorous management. Experiences from Denmark demonstrate the complexity of this task in Chap. 1. Spelling variation, aliases, abbreviations, errors and typos all generate difficulties in further processing and require data cleaning. Similarity measures can be helpful to

identify variants on the fly in further processing, but standardisation of variants of geographical locations, occupations and names—addressed in Chaps. 2, 3 and 4—can make data processing much more efficient, while identifying variants that are not similar at all. Automatic procedures can be helpful for standardisation but generally require expert review.

In the second phase, records that refer to the same person or persons are identified by a process of linkage. Advanced methods for record linkage are reviewed in Chap. 5, with reference to privacy issues that arise when recent data sources are involved. Given the complex reasoning that can underlie genealogical reconstruction, the availability of reconstructions by genealogists in standardised *Gedcom* format can support wider population analyses. The validation and usage of this type of information are discussed in Chap. 6. Whereas family relationships can be deduced from birth, marriage and death certificates from the vital registration or parish registers, the reconstruction of wider social networks may need the analysis of other sources such as notary acts. Multi-source record linkage in this context is addressed in Chap. 7. A comparable complexity was encountered in the challenging project to reconstruct the historical population of Norway, in which data from a wide variety of sources are used. The structure of this process is presented in Chap. 8. Population reconstruction from medieval charters is only possible for the very limited group of people with property worth mentioning in the charters. Probabilistic record linkage on the basis of context information is attempted to arrive at reconstruction in Chap. 9.

In the third and final phase, the information on an individual is combined into the reconstruction of a life course. Whereas record linkage usually focuses on matches between two records or two events, here the full life cycle is taken into account. Catalonia has a unique collection of marriage licences from over 450 years (1451–1905). In Chap. 10, this data collection is analysed to investigate how to utilise this information to reconstruct lifespans in the sixteenth and seventeenth centuries. For many countries, censuses contain key information for population reconstruction, but tracing individuals across censuses over the years is a complex problem. Chapters 11 and 12 report on results using machine learning algorithms for comparisons between nineteenth-century Canadian census records and especially discuss the limitations of the reconstructions and the possible biases, but also the opportunities to study intergenerational social mobility. One way to support the linkage between censuses is the combination with information from the vital registration. An example of such an attempt is described in Chap. 13 for people from the seven parishes on the Isle of Skye and their residence after migration to Scotland. A special population are the 73,000 men, women and children, transported between 1803 and 1853 to the island prison of Van Diemen's Land, now Tasmania, in Australia. The description of the lives of these convicts is discussed in Chap. 14 and encompasses the full process of data collection—including crowd sourcing—linking and life course reconstruction.

The studies and examples in this book originate from a range of countries, each with its own cultural and administrative characteristics, and from medieval charters through historical censuses and vital registration to the modern issue of privacy preservation. Despite all this diversity in place and time, they share the study of the fundamental issues when it comes to model reasoning for population reconstruction and the possibilities and limitations of information technology to support this process.

Gerrit Bloothooft
Peter Christen
Kees Mandemakers
Marijn Schraagen

Contents