

# Introduction: Major Databases with Historical Longitudinal Population Data: Development, Impact and Results

By Sören Edvinsson, Kees Mandemakers and Ken R. Smith

To cite this article: Edvinsson, S., Mandemakers, K., & Smith, K. R. (2023). Introduction. Major Databases with Historical Longitudinal Population Data: Development, Impact and Results. *Historical Life Course Studies*, 13, 186–190. <https://doi.org/10.51964/hlcs14840>

## HISTORICAL LIFE COURSE STUDIES

Major Databases with Historical Longitudinal Population Data:  
Development, Impact and Results

VOLUME 13, SPECIAL ISSUE 4

### GUEST EDITORS

Sören Edvinsson  
Kees Mandemakers  
Ken Smith



## MISSION STATEMENT

# HISTORICAL LIFE COURSE STUDIES

*Historical Life Course Studies* is the electronic journal of the *European Historical Population Samples Network* (EHPS-Net). The journal is the primary publishing outlet for research involved in the conversion of existing European and non-European large historical demographic databases into a common format, the Intermediate Data Structure, and for studies based on these databases. The journal publishes both methodological and substantive research articles.

### Methodological Articles

This section includes methodological articles that describe all forms of data handling involving large historical databases, including extensive descriptions of new or existing databases, syntax, algorithms and extraction programs. Authors are encouraged to share their syntaxes, applications and other forms of software presented in their article, if pertinent, on the EHPS-Net website.

### Research articles

This section includes substantive articles reporting the results of comparative longitudinal studies that are demographic and historical in nature, and that are based on micro-data from large historical databases.

*Historical Life Course Studies* is a no-fee double-blind, peer-reviewed open-access journal supported by the European Science Foundation (ESF, <http://www.esf.org>), the Scientific Research Network of Historical Demography (FWO Flanders, <http://www.historicaldemography.be>) and the International Institute of Social History Amsterdam (IISH, <http://socialhistory.org/>). Manuscripts are reviewed by the editors, members of the editorial and scientific boards, and by external reviewers. All journal content is freely available on the internet at <http://www.ehps-net.eu/journal>.

### Co-Editors-In-Chief:

Joana Maria Pujadas-Mora (Open University of Catalonia & Center for Demographic Studies, Autonomous University of Barcelona) &  
Paul Puschmann (Radboud University)

### Associate Editor:

Eva van der Heijden (Utrecht University)

[hislives@kuleuven.be](mailto:hislives@kuleuven.be)

**The European Science Foundation (ESF)** provides a platform for its Member Organisations to advance science and explore new directions for research at the European level. Established in 1974 as an independent non-governmental organisation, the ESF currently serves 78 Member Organisations across 30 countries. EHPS-Net is an ESF Research Networking Programme.



**The European Historical Population Samples Network (EHPS-net)** brings together scholars to create a common format for databases containing non-aggregated information on persons, families and households. The aim is to form an integrated and joint interface between many European and non-European databases to stimulate comparative research on the micro-level.  
Visit: <http://www.ehps-net.eu>.



HISTORICAL LIFE COURSE STUDIES  
VOLUME 13 (2023), published 30-05-2023

# Introduction: Major Databases with Historical Longitudinal Population Data: Development, Impact and Results

Sören Edvinsson

Centre for Demographic and Ageing Research, Umeå University

Kees Mandemakers

International Institute of Social History, Amsterdam & Erasmus University Rotterdam

Ken R. Smith

University of Utah

## ABSTRACT

Over the last 60 years several major historical databases with reconstructed life courses of large populations spanning decades have been launched. The development of these databases is indicative of considerable investments that have greatly expanded the possibilities for new research within the fields of history, demography, sociology, as well as other disciplines. In this volume spanning seven articles, eight databases are included that have had a wide impact on research in various disciplines. Each database had its own unique genesis that is well described in the articles assembled in this volume. They inform readers about how these databases have changed the course of research in historical demography and related disciplines, how settled findings were challenged or confirmed, and how innovative investigations were launched and implemented. In the end we explore how research with this kind of databases will develop in future.

**Keywords:** Historical demography, Historical microdata, Life course, Social science history, Record linkage, Longitudinal research

e-ISSN: 2352-6343  
PID article: <https://doi.org/10.51964/hlcs14840>  
The article can be downloaded from [here](#).

© 2023, Edvinsson, Mandemakers, Smith

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 AIMS AND CONTENT

Over the last 60 years several major historical databases with reconstructed life courses of large populations have been launched. The development of these databases is indicative of considerable investments that have greatly expanded the possibilities for new research within the fields of history, demography, sociology, as well as other disciplines. At the annual meeting of the Social Science History Association in Montréal in 2017, the session "Development of Major Databases and Their Results From the Beginning Till now" brought together presentations from some of the largest and most well-established databases with life course data, databases that have also been at the forefront of the development in this field. We were well aware in 2017 of numerous additional databases that had been established around the world in recent decades. In his valedictory speech Kees Mandemakers (2023) made an inventory of a total of 54 databases and even this compilation is not exhaustive.

In order to collect, organize, and then publish information on these major databases in a single collection, invitations were first sent to the leaders of about 25 of these databases. We received in most cases positive and enthusiastic reactions and, in case the leaders of a database declined cooperation, it was mostly due to time constraints. We had no specific selection criteria, except that databases had to be actively used and maintained and the primary purpose of the database had to be the (re)construction of individual-level historical life courses. Archived databases, like the Louis Henry dataset (Séguy, 2001) were therefore excluded. Following the first round of invitations, still others joined the collective endeavour, expanding the geographic coverage of our collection. All in all, we are happy to have the assembled contributions representing 24 databases in two special volumes of *Historical Life Course Studies*. The number and diversity of databases represented here is truly impressive!

Our overall strategy of describing these major databases resulted in creating two separate special issues. One, *Content, Design and Structure of Major Databases with Historical Longitudinal Population Data*, edited by George Alter, Kees Mandemakers and Hélène Vézina, deals with the technical and organizational aspects of these databases. They concentrate on aspects such as their origins and evolution including any setbacks, dependence on external funding, content and database designs. The present volume focuses on how the databases contributed to discoveries and insights in historical demography and related fields. Several questions were posed to the leaders of each contributing database: How were previous research questions addressed or altered? What novel lines of inquiry were developed, thanks to the availability of your data? What new knowledge, insights, and scholarly debates were generated that were attributed to the availability of your data? The objective of these central questions was to explore the research productivity and impact of these databases. To summarize the outcome of the investments and labor in a comprehensive way, each contributor was asked to describe the main research contributions resulting from the use of their data, especially with regards to knowledge and insights, that were (1) unavailable before the databases were constructed and (2) could not have been accomplished without having access to individual life course data. In general, the articles assembled in this volume inform readers about how these databases have changed the course of research in historical demography and related disciplines, how settled findings were challenged or confirmed, and how innovative investigations were launched and implemented.

In this volume spanning seven articles, eight databases are included that had a wide impact on research in various disciplines. For databases that are still in early stages of development or for other reasons had limited impact until now, we decided to include information about their impact in the technical articles in the volume mentioned above. This means that the databases included in this impact volume all have their counterpart in the technical one. There is always an exception: the Historical Sample of the Netherlands (HSN), on which the first published article in our volume appeared, does not have a technical counterpart, due to a lack of time. On the other hand, one of the described offshoots of the HSN, the LINKS database, found its own place in the technical volume. The other six databases describing their impact are the Historical Chinese Micro Database, the Demographic Database of Umeå, the Utah Population Database, the Scanian Economic Demographic Database of Lund (SEDD), the Norwegian Historical Micro Database and the Antwerp COR\*- Database. Unfortunately, two major databases that presented a technical article had insufficient time to deliver an impact contribution. These are the two representing Quebec: BALSAC in Chicoutimi and the Programme de Recherche en Démographie Historique (PRDH) in Montreal.

## 2 CONTEXT AND TYPOLOGY OF THE DATABASES

Databases with life course data may be distinguished into databases containing (I) longitudinal data, (II) family reconstitutions and (III) semi-longitudinal data. The differences between I, II and III, is that the latter two categories contain data whereby persons are not reconstructed in a continuous way but on the basis of linked sources, e.g., censuses or, in case of family reconstitutions, only church or civil records. Half of our included databases, HSN, SEDD, DDB Umeå and COR, belong to the 'pure' longitudinal datasets; Utah, Norway and China are of a semi-longitudinal nature; and the LINKS database can be categorized as a family reconstitution dataset.

Each database represented in this volume has its own unique genesis that is well described in the various papers. For example, the launch of the DDB at Umeå University was initially motivated by an interest in the development of literacy. For the Utah Population Database, the impetus was the focus on genealogies, genetics, medicine and family history. At the same time, several common elements and circumstances connect these distinct databases. Their developmental arcs share, for instance, commonalities. Perhaps foremost, most of the databases stand on the shoulders of giants who championed quantitative history and the history of the ordinary person. This includes members of the Cambridge Group for the History of Population and Social Structure (Wrigley, Davies, Oeppen & Schofield, 1997), the Annales School with its advocacy of social history (Séguy, 2016), and the proponents of the life course perspective arguing for the plasticity of human development and the role of history (Kok, 2007). With these intellectual foundations as bedrock, the inevitable influence of technology proved to be a catalyst for accelerating the insights of quantitative history by digitizing archival records and through record linking methodologies that served to reveal the complexities endemic in the reconstruction of human populations.

The birth dates of the distinct databases vary but three in our volume originate from the 1970's when computers and software facilitated data entry, processing and database management. These are the Demographic Database Umeå, the Utah Population Database and the Norwegian Historical Data Centre in Tromsø. Others were the *Registre de la population du Québec ancien* (Université de Montréal) and the BALSAC database (Université du Québec à Chicoutimi). These efforts were followed in the early 1990s with new databases here represented by the SEDD database in Sweden, the Chinese datasets and the HSN database in the Netherlands. More recent work has led to the launch of new databases of which we include the LINKS database and the Antwerp COR\*-database. But most recent databases were actually launched in other parts of the world, including Asia, Australia, and South Africa. Many are described in the technical volume, including a summary of their impact. While the expansion of these infrastructures is impressive and benefits the research community broadly, there remain significant portions of the globe that are not represented, largely due to the lack of archival data and a lack of resources needed to create and maintain complex databases.

Accordingly, of the databases represented in this special issue, the majority are derived from western Europe and North America. Many of these were launched decades ago and have created a legacy through numerous publications, large numbers of trainees, and the development of reliable infrastructures that speak to their stability. From the body of work derived from these databases represented in this volume, we discuss some recurrent themes on the way databases have contributed to the literature, have offered new findings, especially results that would not have been possible without these types of databases. In this overview, we have also introduced some information on findings from the other special issue on the technical aspects of the databases. There we also find more representation of databases from other parts of the world including Korea, Japan, South Africa, Australia, Tasmania, Russia and Spain.

## 3 IMPACT ON RESEARCH

For much of the western world, basic demographic trends and structures were well described and analyzed before the large individual and family level life course databases were developed. But it is worth emphasizing that this is not the case everywhere. More important is that these classic data sources, often highly aggregated, could not address a range of questions about demographic and social mechanisms, such as the role that historic shocks have on individual life course trajectories. To analyze these issues data needs to be collected on a micro level. The databases represented in this

volume have both corroborated past findings while also expanding our understanding in new ways given the richness of the databases and the tools used to analyze them. We will highlight here some of the results for some important aspects of social and demographic history.

*Subpopulations:* While the databases attempt to characterize entire populations in a geographic area, thereby allowing for analyses of all individuals and families, it is clear that the large "sample size" of these databases allows for examining demographic heterogeneity present in all populations. Comparative analyses of the demography for different groups are a common aspect of all the databases represented. A recurrent theme in nearly all the papers relate to differences in racial or ethnic groups, religion, sex, and age.

*History Meets Lives:* A significant advantage of the databases with individual information is that it makes it possible to study how individuals respond to external circumstances, and societal structures and their changes. Shocks may be localized, allowing for comparisons between regions and over time. If you view the dynamics of entire populations as representing what history's video camera has captured (and therefore visible in a database), it is possible to examine intensively and extensively how history shapes lives. Indeed, when intentional changes happen through policy implementation, evaluations of those policies have been conducted.

*Life Course of Individuals and Families:* One of the most important aspects of research using these databases is their life course coverage often from birth to death. A multitude of studies have analyzed different aspects of this, something that was not possible before the development of these databases that cover entire life-spans. This is the case both at individual as well as the family and household level. These dynamics in families and households have been extensively studied as noted in this volume.

*Intergenerational and Familial Studies:* Given the time scale encompassed in these databases, it is possible to see events comprising entire lives for individuals and their ancestors and descendants. As long as included individuals remain in the catchment area covered by the database, it is possible to observe connections, both genetic and social in origin, among relatives as well as those who marry into a lineage. Combining these linkages with information about historical events provides opportunities to see the potential intersection of social history and family history. These opportunities have attracted not only historical demographers but also geneticists and evolutionary biologists.

*Comparative Analyses:* The growing number of large historical databases available to the research community create the opportunity for examining common research questions across social and historical contexts. Given the heterogeneity of decades and locations embedded in these historical databases, it is now possible more than ever to expand on comparative analyses that involve a larger number of populations, all of which serve to improve our understanding of social and family history.

*Database Expansion:* Given the success and productivity of the large historical databases included here, it is reasonable to consider adding to this portfolio by building new resources from previously omitted areas of the world. Offering expertise and direction to investigators who represent previously underdeveloped historical populations from those who have already done it can prove to be an effective way to improve data coverage.

## 4 FUTURE DIRECTIONS

As we have introduced this volume on history and demography and key databases, we have likewise adopted a more historical perspective about their intellectual and scientific impact — that is, what these databases have done. Here we briefly consider what these databases could do going forward. The existing databases saw their births in a time where the central records were decidedly amenable to the social sciences. This meant that archival, census and religious sources and the demographic, family, and spatial information they contained formed the essential ingredients needed in the database recipe. With this as foundation, we are now imagining how to build off of it. Several new opportunities now (and in some cases have been) present themselves. First, the advent of population genetics means that the growing number of countries that collect genetic material may be used to link to historical databases, assuming the appropriate human subject protections are in place. One can imagine more



collaborations between historical demographers and geneticists studying disease risk as well as demographic phenomena like fertility and mortality using DNA information. Indeed, knowing genetic variants in contemporary populations, when joined with family data in the databases, may allow one to infer what these variants are in past populations (Adams, Lam, Hermalin, & Smouse, 1990).

Second, the rise of geographic information systems (GIS) has paralleled the rise of big genetics and in like fashion, offer new directions that build on the existing historical databases. Since the databases described in this volume generally have geographic information within them, it is now possible to link, at a spatial level, additional data to individuals and families. An excellent example of this are the data available in the IPUMS National Historical Geographic Information System (NHGIS) in the US. But in general, joining these GIS variables can only enrich the research opportunities that will provide novel insights into our understanding of historical populations.

Finally, it is possible to consider how to do comparative analyses in new ways. Specifically, the comparison can now be done by actually connecting people and lineages in one population to another. In the spirit of linking European and North American databases, it is possible to observe and analyze people who left one country for another in these databases and compare them to people who did not emigrate. So, in addition to studying two independent populations, it is conceivable to see the same person or their relatives and descendants in both countries.

## REFERENCES

- Adams, J., Lam, D. A., Hermalin, A. I. & Smouse, P. E. (1990). *Convergent issues in genetics and demography*. New York: Oxford University Press.
- Kok, J. (2007). Principles and prospects of the life course paradigm. *Annales de démographie historique*, 113(1), 203–230. doi: [10.3917/adh.113.0203](https://doi.org/10.3917/adh.113.0203)
- Mandemakers, K. (2023, January 20). “You really got me”. *Ontwikkeling en toekomst van historische databestanden met microdata* [Development and future of historical databases with microdata] (Valedictory speech). Erasmus University, Rotterdam, the Netherlands. doi: [10.25397/eur.23256467.v1](https://doi.org/10.25397/eur.23256467.v1)
- Séguy, I. (2001). *La population de la France de 1670 à 1829: l'Enquête Louis Henry et ses données* [The population of France from 1670 to 1829: The Louis Henry survey and its data]. Paris: INED.
- Séguy, I. (2016). The French school of historical demography (1950–2000). In: A. Fauve-Chamoux, I. Bolovan, & S. Sogner (Eds.). *A global history of historical demography. Half a century of interdisciplinarity* (pp. 257–276). Bern: Peter Lang.
- Wrigley, E. A., Davies, R. S., Oeppen, J. E., & Schofield, R.S. (1997). *English population history from family reconstitution 1580–1837*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511660344](https://doi.org/10.1017/CBO9780511660344)