

LONGPOP and IDS. Personal Reflections on our Collaboration With Kees Mandemakers

By Sam Jenkinson, Hideko Matsuo and Koen Matthijs

To cite this article: Jenkinson, S., Matsuo, H. & Matthijs, K. (2021). LONGPOP and IDS. Personal Reflections on our Collaboration With Kees Mandemakers. *Historical Life Course Studies*, 10, 81-85. <https://doi.org/10.51964/hlcs9572>

HISTORICAL LIFE COURSE STUDIES

Not Like Everybody Else.
Essays in Honor of Kees Mandemakers

VOLUME 10, SPECIAL ISSUE 3
2021

GUEST EDITORS
Hilde Bras
Jan Kok
Richard L. Zijdeman



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 openjournals 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 <https://openjournals.nl/index.php/hlcs>.

Co-Editors-In-Chief:

Paul Puschmann (Radboud University) & Luciana Quaranta (Lund University)
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>.



LONGPOP and IDS

Personal Reflections on our Collaboration With Kees Mandemakers

Sam Jenkinson

KU Leuven, Belgium

Hideko Matsuo

KU Leuven, Belgium

Koen Matthijs

KU Leuven, Belgium

ABSTRACT

The Leuven research team working on historical demography is grateful for their opportunity to have collaboratively and intensively worked with Professor Dr. Kees Mandemakers over an extended period of time. We wish him a wonderful emeritus status, not only in academia, but also in the warm nest of his family, relatives, children, and grandchildren. The three of us have known Kees for some time, but most closely since 2014, when we became formally engaged as project partners under the so-called LONGPOP-project, an EU funded Marie Curie grant, named Methodologies and Data mining techniques for the analysis of Big Data based on Longitudinal Population and Epidemiological Registers. The importance of our close professional relationship is best demonstrated by our work in producing the COR*-IDS 2020 database. The historical demographic dataset for the Antwerp arrondissement, a letter sample COR*-2010, recorded total sample size of +/- 33,000 residents of Antwerp for nearly seven decades and was already available. The LONGPOP project began in Autumn 2016, setting in motion a collaboration between ourselves at KU Leuven and Kees and his colleagues at the IISH. From the outset, we had purposefully worked closely with Kees' team, utilising their premier expertise in database management and the IDS towards the new release of our database, the Antwerp COR*-IDS dataset. Here we set out our recollections of that intellectual process, encompassing the personal and professional reflections of our close working relationship.

Keywords: Historical demography, Historical database management, Intermediate Data Structure (IDS)

e-ISSN: 2352-6343

DOI article: <https://doi.org/10.51964/hlcs9572>

The article can be downloaded from [here](#).

© 2021, Jenkinson, Matsuo, Matthijs

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 PERSONAL PROLOGUE

Dear Kees, you have begun an exodus, the transition from an active to a slightly less active academic life. Others are now helping you to make that transition bearable, for example by organising a few rites of passage. This special issue is one such rite.

The Leuven research team working on historical demography is grateful for their opportunity to have collaboratively and intensively been working with you over an extended period. You always insisted on the need for exact registration and the methodological precision necessary for successful data linkage. And you have certainly done a lot of linking and standardizing of data in recent years, both nationally and internationally. By stubbornly maintaining what a correct approach to the so-called Intermediate Data Structure (IDS) should look like, you have set the bar high. In the long run, that will benefit everyone.

The Leuven research group wishes you a wonderful emeritus status, not only in academia, but also in the warm nest of your family, relatives, children, and grandchildren. You will certainly remain associated with research institutions for a long time to come and will continue to engage with important academic issues. After all, you will be tremendously highly remembered and regarded there. Do not let go of that, keep doing that, and keep sharing it with us.

Although Koen has known Kees for decades in the context of, in particular, the Leuven-initiated historical demography meetings, let us take an overview of the academic trail we have taken together intensively over the past decade. We are proud of this body of work, and we hope that the following generations of researchers will continue to build on it.

On behalf of all the Leuven colleagues, we thank you for your dedication, your achievements and your inspiring insights.

2 PROJECT FRAMEWORK: LONGPOP

The three of us have known Kees for some time, but most closely since 2014, when we became formally engaged as project partners under (the preparation of) the so-called LONGPOP-project, an EU funded Marie Curie grant, named *Methodologies and Data mining techniques for the analysis of Big Data based on Longitudinal Population and Epidemiological Registers* (<http://longpop-itn.eu/>). This international project involved eleven institutions, including universities (Leuven amongst them), research institutes, enterprises and public administration from seven European countries and the USA. The Spanish National Research Council (CSIC-CCHS-IEGD) coordinated the overall effort.

3 IDS AND COR*-SAMPLE

The importance of our close professional relationship is best demonstrated by our work in producing the COR*-IDS 2020 database (Jenkinson, Anguita, Paiva, Matsuo, & Matthijs, 2020). The historical demographic data set for the Antwerp arrondissement a letter sample COR*2010 recorded total sample size of +/- 33,000 residents of Antwerp for nearly seven decades and was already available (Matthijs & Moreels, 2010). This data set was collected from historical records of population registers and vital registration records covering births, marriages, in/external migrations and deaths. However, it had been some time since a new version of the database had been released. The Leuven LONGPOP-subproject, in partnership with Kees and the International Institute of Social History (IISH), represented a unique and valuable opportunity for this update, and allowed us to exploit the state-of-the-art expertise and knowledge that Kees and IISH possess.

One of the most important aspects of this knowledge was Kees' work on developing the concept of the Intermediate Data Structure (IDS). This work was executed, in collaboration with George Alter, over many years. The primary aim of IDS is to facilitate the use of historical data sets for cross-national

comparative analysis by creating one single format for data storage and extraction. The quality and quantity of information in many historical datasets is amazing, but their individuality and complexity limits their applicability to cross-national comparative analysis. Each is painstakingly created by their respective database owners, incorporating diverse methodological and conceptual practices. This means that the differences between them, in terms of how the data is specified and stored can be numerous and complex.

We consider the IDS to be an important, indeed revolutionary development in the field of historical demography. This conclusion flows from our own research experiences, which have shown that working with raw historical datasets requires countless hours of complex data preparation, especially if using multiple datasets for comparative cross-national analysis. The IDS overcomes these problems by advancing one common protocol for storing, specifying and extracting data for historical populations. Prior to LONGPOP, Kees had been working continuously with the European Historical Population Sample Network (EHPS-Net, <https://ehps-net.eu/>). Leuven was part of this international consortium, and we choose to focus on re-developing the Antwerp COR*2010 database, because of the many possibilities that it would open for the international research community. These possibilities arise in part from the combination of vital registration records of births, marriages and deaths, and the population registers that were collected via a letter-based sampling method using COR*-names. Of special note is the fact that the database contains information on both international and interprovincial migrants residing in the Antwerp arrondissement in the period of the early 18th to the early 19th century. The Antwerp COR*-IDS database 2020 is therefore a transformed and harmonized demographic database in a cross-nationally comparable format which was mostly financed by LONGPOP and prepared for the special issue of 'Content, Design and Structure of Major Databases with Historical Longitudinal Population Data' in *Historical Life Course Studies* (2020).

4 IDS PRODUCTION PROCESS: KEES AND US

The LONGPOP project began in Autumn 2016, setting in motion a collaboration between ourselves at KU Leuven and Kees at the IISH. From the outset, we had purposefully proposed to work closely with Kees' team, utilising their premier expertise in database management and the IDS towards the new release of our database, the Antwerp COR*-IDS dataset.

Ultimately, this process would involve an intense travel schedule between our institutions, many online video calls and more emails. Kees stayed deeply engaged throughout: not just on an abstract theoretical level but immersed in the nitty-gritty details. Kees never seems to be happier than when given the opportunity to explain his work to someone else, or to discuss in detail the technical issues relating to historical databases. After the LONGPOP project came into place in fall 2016, we had a highly instructive and consequential meeting with Kees in spring 2017. At that time, Kees had already identified a number of technical, yet crucial issues of the production process for COR*-IDS. Some examples included the reviewing of relationship variables arising from the population register; individual and intergenerational linkages; applying Levenshtein distance measures; correction and standardization of information on names, gender, and dates across sources; quality of records concerning key variables, such as households and occupations; and time-stamp information. Throughout this, Kees emphasized: 'It is not the data that is important to look into, but the original source!'. In summer 2017, after our subsequent visit to Amsterdam, the nitty-gritty evaluation of the COR*-sample was initiated, and by summer 2018, the first part of our work in evaluating and standardizing variables, and most importantly, reproducing the individual and intergenerational linkages to evaluate the original work of COR*-sample, was finalized.

5 EXCHANGE OF VISITS: LEUVEN AND AMSTERDAM BROADENING COLLABORATIONS

While in-house work on COR*-IDS production continued throughout 2018, the collaboration between Leuven and Amsterdam was strengthened by exchanges of two researchers. Firstly, Francisco Anguita, Kees' colleague, spent 5 weeks in Leuven (October and November 2018), and secondly this was followed by Sam's 4 week stay in Amsterdam. While these visits were initiated by the Marie Curie programme, these exchanges were an exceptionally valuable period for all of us. Our collaborations were grounded in many discussions in and outside of regular working hours. Francisco developed the entire algorithm to create familial relationships based on population registers. Working with Kees, he had developed an expertise in linking individuals from different sources and even across continents: he was linking Dutch emigrants in HSN and family members to the US, recorded in the Census. One particularly fascinating aspect of this linking involved the numerous linguistic patterns through which mainly Dutch names had been changed, such as, Hendrickx to Henry. With Francisco's algorithm for creating familial relationships within households, we managed to establish an additional one-third of linkages of parents and children that became inputs of individual-individual tables for COR*-IDS. Francisco's stay concluded with three of us, Sam, Francisco and Hideko, visiting the State Archive in Antwerp-Beveren, where we asked the head of the state archive, Dr. Johan Dambruyne, to let us review volumes of population registers. In doing so, we paid close attention to something Kees has always told us was important: 'you must always look at the original source'. We flipped the pages of the population register, containing lots of information on household characteristics, reading through heads and familial relationships across different years. Thanks to the access made by Dr. Dambruyne, our visit became a memorable expedition, ending with a fine 'pintje' at the station cafe and a feeling of confidence in what we had accomplished.

As noted, Sam went to Amsterdam in the winter of 2018. Diogo Paiva, who was also working at IISH on the geographical information of HSN, made a valuable contribution to COR*-IDS production through his expertise of standardization of addresses allocating geographical information (i.e. XY coordinates). Kees was also an active partner in this process. In the summer and fall of 2019, he was present in our presentation in Pécs at the European Society of Historical Demography conference, and our final project meeting of LONGPOP in Edinburgh to support and ensure that the goals of COR*-IDS would be fully achieved and the research output valorised.

In December 2018, Sam went to the IISH in Amsterdam. We had chosen the IISH due to their expertise in a number of strategically crucial areas, and we believed this collaboration would be highly beneficial, indeed crucial, to our work in Leuven and Sam's contribution in particular. The IISH expertise included the handling and developing of historical databases, knowledge of historical primary sources and expertise in transforming historical data into IDS. In each of these areas, Kees and his team are the foremost experts within the field of historical demography. The placement would involve moving to Amsterdam for roughly 4 weeks in December 2018. As wonderful a city as Amsterdam is, this period was bitterly cold and windy. A different climate altogether from what Sam found in Leuven and his team. Everybody was warm, welcoming, and patient with Sam. Kees has truly created a wonderful team of people who work closely together, take an interest in each other, and welcome new people with open arms. This is clearly seen in small gestures such as how they check in each morning, have lunch together, and always stop by each other's offices to say good night before leaving.

Kees frequently joined his team for lunch, another sign of just how close and friendly they all are. Sam still remembers characteristically Dutch composition of most people's lunches. Always a croquette, a bread roll, some sliced cheese and definitely a carton of milk ('karnemelk', which is a buttermilk to be precise), which frankly, is all that you need and remain a habit for Sam even when he was back in Belgium.

The most important positive reflection we have about the way Kees works, is that he really sees the value in going back to basics and knowing the essentials in detail. For him, it is vital that anyone working with historical demographic data first gets to know their primary historical sources in detail. You may be able to manipulate the data in all manner of sophisticated ways, but if you do not know the 'what, where and how' about how this data came to exist and come into your hands, then you will not truly know what you are doing and will likely risk running into basic mistakes.

With this in mind, the first week in Amsterdam involved looking through copies of historical Dutch population registers in order to decipher their contents and input them into the system. To some, this may sound dull, or even unnecessary, but Sam felt it a valuable research experience. It taught him much about his own historical sources, their limits, how their interpretation could be improved, and how the nature of what they capture can change dramatically over time.

Later on, the work would move into software testing of an IDS database and looking for errors, which could highlight any mistakes in the process of assembling the database. Although it was apparent earlier, it became increasingly obvious that Kees is someone who really enjoys the nitty-gritty details of his work and loves nothing more than taking the time to patiently explain and teach them to someone new. He seems to be someone who enjoys not only the high-level discussions typical of a professor, but equally the particulars and technical intricacies of his work.

6 WORKING IN BIG CHAOS: COR*-IDS DURING THE COVID-19 CRISIS

In 2020, during the COVID-19 crisis, our collaboration was intensified. We needed to finalize a pending journal article and most importantly prepare our database for release. Our online video meetings continued during the lock-down period since March, when evaluating the familial relationships and elaboration on core concepts such as occupations and its associated time stamp information. These meetings became frequent as the Leuven-Amsterdam collaboration was reconstituted from everyone's home residence in Lisbon, Ghent, Brussels and Amsterdam. These online video meetings produced memorable experiences as they migrated from one room to another or sometimes from the bathroom when a young child wakened from her nap perhaps from hearing so many unfamiliar voices.

The journey of COR*-IDS production is now in the last stage after several resubmissions and reviews from George and Kees. 2020 has been the year of our most intensive interactions with Kees. We do not know how much that has to do with COVID-19, but Kees' enthusiasm and persistence on IDS was a major factor. His scientific approach is thorough and rigorous. His keen eye for details brought many valuable suggestions for our work including specific figures and tables that added clarity and substantive value. We have the utmost respect for his intellectual enthusiasm and scholarly standards, and realize how privileged it was to work with him. And finally our collaborations have played their role in the production of COR*-IDS, setting one milestone to the following generations of researchers.

REFERENCES

- Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal microdata, version 4. *Historical Life Course Studies*, 1(1), 1–26. Retrieved from <http://hdl.handle.net/10622/23526343-2014-0001?locatt=view:master>
- EHPS-Net (n.d.). Website: <https://ehps-net.eu/>.
- Jenkinson, S., Anguita, F., Paiva, D., Matsuo, H., & Matthijs, K. (2020). The 2020 IDS release of the Antwerp COR*-database: Evaluation, development and transformation of a pre-existing database. *Historical Life Course Studies, Online first*. Retrieved from <https://hdl.handle.net/10622/23526343-2020-0010>
- LONGPOP (n.d.). Website: <http://longpop-itn.eu/>.
- Matthijs, K., & Moreels, S. (2010). The Antwerp COR*-database: A unique Flemish source for historical-demographic research. *The History of the Family*, 15(1), 109–115. doi: [10.1016/j.hisfam.2010.01.002](https://doi.org/10.1016/j.hisfam.2010.01.002)