

# The Groningen Integral History Cohort Database. Development, Design and Output

By Richard Paping and Dinos Sevdalakis

To cite this article: Paping, R., & Sevdalakis, D. (2022). The Groningen Integral History Cohort Database. Development, Design and Output. *Historical Life Course Studies*, 12, 78–98. <https://doi.org/10.51964/hlcs12033>

## HISTORICAL LIFE COURSE STUDIES

Content, Design and Structure of Major Databases with  
Historical Longitudinal Population Data

VOLUME 12, SPECIAL ISSUE 5,  
2020

### GUEST EDITORS

George Alter  
Kees Mandemakers  
Hélène Vézina



## 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 [hlcs.nl](http://hlcs.nl).

### Co-Editors-In-Chief:

Paul Puschmann (Radboud University) & Luciana Quaranta (Lund 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 12 (2022), published 20-06-2022

# The Groningen Integral History Cohort Database

## Development, Design and Output

Richard Paping

University of Groningen

Dinos Sevdalakis

University of Groningen

### ABSTRACT

The Groningen Integral History project launched in 1987 aimed to sketch the lives of people from various social classes in the Dutch province of Groningen in the 19th and early 20th century. One part was the creation of the Groningen Integral History Cohort Database (GIHCD), reconstructing complete individual life courses of 5,280 persons (RPs) born between 1811 and 1872. The quality of the database has become very high by now, despite the lengthy and difficult process of shaping it over 35 years. More than 98% of the RPs (and for some parts of the database even more than 99%) could be followed until their death or until a migration abroad. Even for the life courses of those moving abroad information is available for most RPs. In this article, we primarily focus on the rural part of the database (n = 4,320), the quality of which is the highest and has had the most significant tangible research impact. Building on information from the Dutch civil registration system (from 1811) and the population registers (from 1850), the database includes multiple individual-level variables. In the technical part of the article, we provide an extensive overview of the available variables and summarize the transformation of the rural part of the database into an Intermediate Data Structure (IDS). Since the early 1990s, historians from the University of Groningen have used GIHCD in quite some publications. At the end of this article, we provide a summary of the main outcomes of these publications.

**Keywords:** Demography, Historical databases, Life courses, Groningen, Dutch Civil registration, IDS

e-ISSN: 2352-6343

DOI article: <https://doi.org/10.51964/hlcs12033>

© 2022, Paping, Sevdalakis

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/), 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 INTRODUCTION

The Groningen Integral History Cohort Database (GIHCD) consists of a regional sample of more than 5,000 life courses of persons born between 1811 and 1872 in the northern Dutch province of Groningen. This province consists of a large city (Groningen) and its surrounding countryside with numerous smaller and larger settlements organised in about 60 municipalities in the 19th century. Most of the individuals involved were followed during their life throughout the Netherlands, and largely also abroad until well into the first half of the 20th century. The database's main strength is its excellent quality, as a very high percentage of the included persons could indeed be followed throughout their entire life course.

This article first presents an overview of the history of the Integral History Project out of which the GIHCD developed from 1987 onwards. Secondly, we describe the rather unsystematic way in which the database was constructed over the years, partly in response to changing research questions. In the next section, we discuss what the database looks like at the moment, presenting its content and structure. In the last part, we will present an overview of the publications that made use of the data stored in the database. In general, the article also provides an illustration of how technological developments over more than thirty years, in addition to a continuous lack of time and money, can shape a database like GIHCD into its present form.<sup>1</sup>

## 2 THE ROOTS OF THE DATABASE

### 2.1 THE START OF THE INTEGRAL HISTORY PROJECT

In 1987, researchers from both the University of Groningen (Pim Kooij and Marten Buist) and the University of Utrecht (Gerard Trienekens and Theo van Tijn), with support from the organisation for historical research Stichting voor Historisch Onderzoek (SHO), one of the predecessors of the government-funded Dutch organisation for scientific research NWO, launched the Integral History Project (Kooij, 1993a, p. 2). The primary goal of the project that led to the creation of the GIHCD, was to write an integral history of different regions in the north and the south of the Netherlands. Through this approach the initiators sought to solve the ongoing fragmentation of historical research by operationalising the concept of "quality of life" (Kooij & Sleebe, 1991; Trienekens, 1987, 1993). In short, the theoretical idea was to compare the aims of ordinary people regarding various aspects of their lives with their perception of what was happening in "reality", thus enabling historians to research the quality of life of people in the past. According to this reasoning, major differences between ambitions and perceptions of people will have led to social tensions that would clearly show up in the sources. As general developments — for instance technological and economic progress, modernisation in other respects and democratisation — influenced these aims and perceptions, this would also result in opportunities to connect data on micro and macro level.

Although at first sight events were not the core of the Integral History Project, the main way to trace indications of changes in ambitions and perceptions was by doing extensive and detailed archival research on specific individuals and communities representing different societies. Four research pillars were defined, that had to result in a similar number of large databases: 1. Cohort analysis: research on the life course of the inhabitants of selected municipalities; 2. Structural analysis: research based on snapshots of the household structure of the inhabitants in these selected municipalities taken every 20 years; 3. Financial analysis: research on the developments in the provincial and municipal expenditures using the annual accounts (Duijvendak & Blijham, 1994) to study the political priorities of the local and provincial elites; 4. Opinion analysis: research based on the local newspapers to find indications of tensions in society. For this article, we concentrate on pillar 1 and to some extent on pillar 2, because they at least have had some long-lasting research output.

Originally, two Dutch cities and their surrounding regions were selected to be studied in the period 1770–1914. On the one hand, the city of Groningen and the clay soil region situated largely north and east of it, and on the other hand, eastern Northern Brabant with the capital of Den Bosch (van Oudheusden & Trienekens, 1993). So, next to the capital, nine more or less rural municipalities were chosen to be investigated in detail for both regions.

Regarding the Northern Brabant part of the project, it must be noted that it was attempted to reconstruct the life courses (pillar 1) engaging many volunteers. However, this data collection — partly due to unsolved

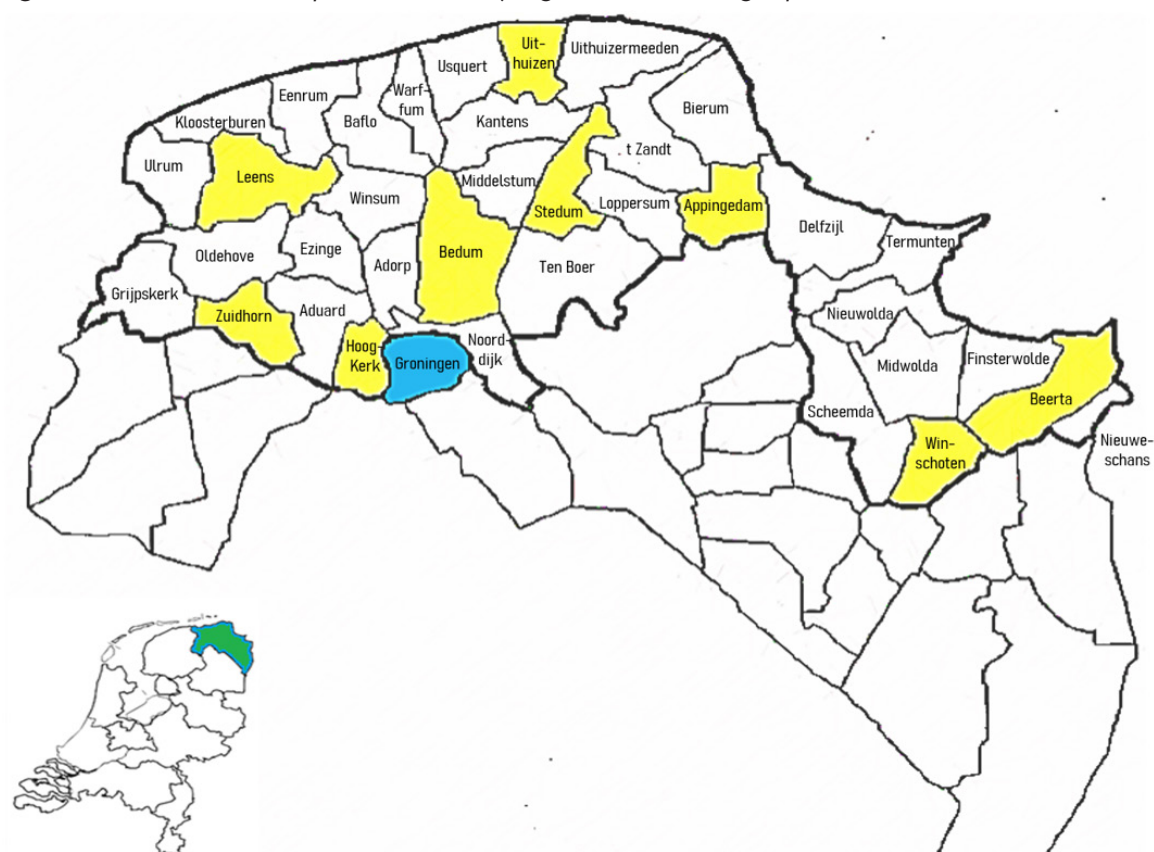
1 The authors want to thank Geurt Collenteur, Maarten Duijvendak, Frouke Hansum, Kees Mandemakers and H el ene V ezina for their many useful comments.

technical problems — did not result into a database that could be analysed easily. Still, some results from the Brabant cohort analysis were published in Schrover (1998). Largely due to the different sample years chosen for each municipality, the collection of the household structures (pillar 2) turned out to be problematic as well. Preliminary results were presented by Trienekens (2004), while a database called "Het Huishouden in Oostelijk Noord-Brabant 1800–1920" is available.<sup>2</sup> However, the overall output of the Northern Brabant part of the project is limited. Consequently, this article will concentrate on the Groningen part of the project.

The Groningen clay region was largely defined as those municipalities whose surface area was covered by clay for more than 50% (Paping, 1995, pp. 18–21). The selected region covers more than half the province including 36 out of the 57 19th-century municipalities (see figure 1). The nine municipalities were chosen in such a way to ensure an even geographical spread over the region. They were largely rural, though also deliberately comprised of the small old town of Appingedam with its adjoining rural villages and the somewhat larger semi-urbanised settlement of Winschoten. However, for the sake of convenience, we will call these nine municipalities rural in opposition to the large city of Groningen.

The Dutch scientific fund SHO, in addition to financing PhD's, provided the Integral History Project in 1987 with some funding to develop the four databases. In the Groningen part of the project, this money was used partly to finance a fulltime research assistant, who constructed a digital database of government accounts (pillar 3). However, the amount of funds supplied by SHO were by no means sufficient to finance the construction of the four large historical databases, just mentioned. The University of Groningen also funded additional PhD's and student-assistants to participate in the project.

Figure 1 *The municipalities of the clay region in the Groningen province*



Source: Paping (1995).

Explanation: The so-called 'clay area' is the area surrounded by the thick line and the named municipalities. The municipalities in yellow and blue are the ones that are included in the GHICD database.

2 See <https://www.narcis.nl/research/RecordID/OND1296249/Language/nl>

## 2.2 THE STRUCTURAL ANALYSIS DATABASE

The aim of the structural analysis was to take snapshots of the household situation every 20 years from 1815 onwards as benchmarks to study changes over time. The transcription and summarising of the population registers of the nine selected rural municipalities was done by several persons working on PhD-projects, some supported by student-assistants. Originally, a similar endeavour was planned for the city of Groningen for the period before 1870, as the years 1870–1910 had already been covered in the thesis of Kooij (1987). However, the urban part of the project was never completed.

The structural analysis was confronted with three major problems. Firstly, there was a lack of sources. The intended timespan of the project was 1770–1920; however, population registers or micro census-data were missing for the older period. Therefore, it was decided to restrict the data collection to the census years 1815, 1829/1830, 1850, 1870, 1890 and 1910, reconstructing the household situation around January 1st of those years. Unfortunately, for most of the selected municipalities the (census) registers were missing for 1815 and 1830. From 1850 onwards, nationwide introduced dynamic population registers were used, except two municipalities for which these registers were missing for the first decade. Secondly, it often proved difficult to reconstruct the exact household situation in 1870, 1890 and 1910 as in the countryside many of these dynamic population registers cover periods of twenty years or more, for instance from 1860 to 1900, while the original ten-yearly census forms had been destroyed. This made it very difficult to ascertain which persons (including live-in servants) lived within the households at the measure points from 1870 onwards. Thirdly, the reliance on only a few PhD-candidates — due to the lack of funding — meant that the years 1870, 1890 and 1910 were only done to a limited extent. Consequently, only for 1830 and 1850 a digital database on household structures is available for a substantial number of municipalities (see table 1).

As laptops and the like were still rare at the end of the 1980s, all households were registered on a paper form. Standardly, all data of the head of the household and eventual partners were transcribed including the names. For the other members of the household some data were transcribed as well, though in much less detail than available in the registers in an attempt to reduce the time needed to collect the data. After filling out all the forms, the information was translated into numbers using an extensive codebook (with numbers indicating, for instance, household types, occupations, and so on). These numbers were digitalised on tape and analysed by way of SPSS using a mainframe computer. Around 1995, a LOTUS 1-2-3 spreadsheet database was created (later on converted into EXCEL) with a household on every row. The numerical codes were reconverted into the corresponding meaning in text. Next, some of the not yet digitalised information from the still preserved original forms was added to the database. An important feature of this dataset is that for more than half of the municipalities in 1830 and 1850 information on the amount paid in the local taxes could be added (the so-called 'hoofdelijke omslag', a tax based partly on income and partly on wealth: Kooij, 1993b, pp. 145–155; Paping, 2010; Voerman, 2001, pp. 267–272).

Table 1 *Overview of the available databases regarding household structure constructed for the nine rural municipalities of the Groningen Integral History Project*

	1830	1850	1870	1890	1910
Appingedam	Missing	C/D	C	C	C
Bedum	C/D/T	C (partly)	C/D/T	C/D	C/D
Beerta					
Hoogkerk	Missing	C/D/T	C/D/T	C/D/T	C/D/T
Leens	C/D/T	Missing			
Stedum	C/D	C/D			
Uithuizen	Missing	C/D/T	C	C	C
Winschoten	C/D	C/D/T	C/T	C/T	C/T
Zuidhorn	Missing	C/D/T		C/D	C/D

*Explanation: C: Reconstructed snapshots; D: Digitalised and standardised databases; T: Tax information available. Empty cells: sources are available, though no databases have been reconstructed.*

The structural analysis databases were used for several publications: Voerman (2001) made an in-depth analysis for Winschoten in combination with extra data on the mixed rural-urban municipalities of Hoogezand and Veendam in the eastern Groningen peat districts neighbouring the Groningen clay soil region. Paping and Collenteur (1998) used it for the long-term development of the occupational structure of household heads until 1910. Paping (2010) researched the relation between occupation, social position and tax performance within households in the period 1830–1850; and lastly Paping (2018) analysed the household structure using the same data.

## 2.3 THE CONSTRUCTION OF THE GIHCD

This article focuses in particular on the most prominent database constructed within the Groningen Integral History Project: the cohort database with individual life courses from ten municipalities (including the city of Groningen, see figure 1), selecting 120 births from cohorts starting with a 20-yearly interval (1811, 1830, 1850, 1870 and 1910). Although the original aim included the birth cohorts of 1770 and 1790, it soon became clear that this was not feasible. For the three villages later constituting the municipality of Hoogkerk, the first 120 baptisms for the cohort of 1770 and 1790 were collected by Vincent Sleebe (1993). However, it proved extremely difficult to find any other information about these persons. In Groningen, the large majority of the families did not have a surname before 1811, making it difficult to trace persons outside the village without extensive genealogical research. Initially, it was hoped that existing registrations of protestant church members (*lidmaten*) would indicate migrations, but these registers proved scarce or incomplete and covered only a minority of the population. This was even more detrimental, as the naïve initial research assumption that migration was very rare in the countryside proved to be completely beyond reality. How to deal with migrating cohort members became a persistent issue in the project as we will show later on.

The project originally planned a series of books on all 10 municipalities involved starting with Hoogkerk, a tiny municipality right next to the city of Groningen. This characterised the optimistic atmosphere in the first years of the project: ultimately only the book on Hoogkerk (Kooij, 1993) was published. The contribution to this book that makes extensive use of the GIHCD clearly shows the huge problems with which the construction of the cohort analysis was confronted (Clement, 1993). Birth cohorts were constructed for 1811, 1830, 1850 and 1870. As the population of Hoogkerk was very limited, it sometimes took even six years to achieve a cohort of 120 consecutive births. The cohort of 1811 started in August, with the beginning of the official Dutch civil registration. The birth certificates offered much more information (occupations and ages of people involved) than the baptism registers which had to be used before August 1811. At first, children were only traced within the municipality of birth and in the city of Groningen. As most of the Hoogkerk children disappeared without a trace, this search was expanded to neighbouring rural municipalities like Aduard, where indeed a few of the Hoogkerk cohort members were found.

In the 1980s, it was still difficult to track down a specific person in the civil registration. Only 10-yearly indices of births, deaths and marriages existed for each municipality, supplying merely the registration date of the certificates. As it was not uncommon to have several people with the same name, all hits had to be checked one by one. Fortunately, this could be done at the provincial archive, where an increasing amount of films could be found containing the civil registration of the provincial municipalities before 1900. A notable exception were the records of the city of Groningen, which were kept in the city archive. In 2002 the two archives fused (Groninger Archieven) and the work could be concentrated in one building.

Originally, the core of the database was formed by standardised paper forms for each Research Person (RP). These forms had room for information about: 1. The parents (occupation, birth date, birth place, ability to sign birth or marriage certificate); 2. Marriage (date and location) and characteristics of the partner (occupation, birth date, birth place, ability to sign marriage certificate); 3. Children (birth date, occupation and moment they left the parental household, including stillbirths); 4. The religion, the household situation and exact place where the RP lived according to the census (1815, 1830 and 1840) or the population register (from 1850 onwards); 5. Migration dates and destinations. Extra forms could be added when necessary. Consequently, the original database contains information on three generations, not only on the RP, but also on his or her parents, on marriage partners and on the place and date of birth of children.

The advantage of the paper system was its flexibility, as it was easy to include interesting additional information found in archives. The disadvantages were, firstly, that it was often unclear what the source of the transcribed information was and, secondly, that forms were being filled in very inconsistently. The form gave a lot of decision power to the individual researcher about which information to be included. Consequently, a lot of data was not transcribed, making it often necessary to go back to the original sources and replenish the data

on the forms. Most of the forms were filled in by student assistants and research assistants, though also sometimes by history students as part of a paper they wrote during a research course under the supervision of project leader Pim Kooij. Kooij himself did most of the cohorts of the city of Groningen. The handwriting, and also the colour of the pen, pencil or marker provide indications on who originally transcribed the data. Over the course of time, many different researchers worked on one individual RP.

Before 1995, the tracing process in the civil registration proved extremely difficult and time consuming. Especially the high number of persons in the province of Groningen sharing the same name made it time-consuming to establish if a certificate belonged to the RP in question or not, without reading the certificate itself. In 1993/1994 all funds of the Integral History Project had been used. All life courses from the cohorts from 1811 until 1870 had been constructed, except migration routes outside the birthplace. A few student-assistants — financed by the Groningen Faculty of Economics — had tried to improve this; however, it was too much work. In this phase, RPs were followed in the municipalities in the clay parts of the province, and not only in the city of Groningen. During this period, project leader Kooij created a REFLEX-database with a structured summary of the assembled data of every individual based on the paper forms. In Figure 2, we sketch the technical transformations the database has undergone since 1987.

The Integral History Project received some fresh money from the Dutch research foundation (NWO) to stimulate research cooperation with historians from the former USSR. Since the end of the Communist regime, there was a rising interest in Western research methods among Russian historians, together with a change of focus from political history (mainly the history of communism) to the history of common people. This Dutch-Russian cooperation resulted in several workshops in both Russia and the Netherlands and in two volumes: *Where the Twain meet* (Kooij, 1998) and *Where the Twain Meet Again* (Kooij & Paping, 2004). Several Russian contributions applied the Groningen method of cohort analysis (Akolzina et al., 2004; Dyatschkov et al., 1998; Golubeva, 1998; Shustrova & Sinitsyna, 2004; Sinitsyna, 1998). However, source problems (serious gaps in birth, marriage and death records and difficulties with record linkage) were even larger for 19th century Russia, resulting in databases containing only a limited number of individuals with substantial information.

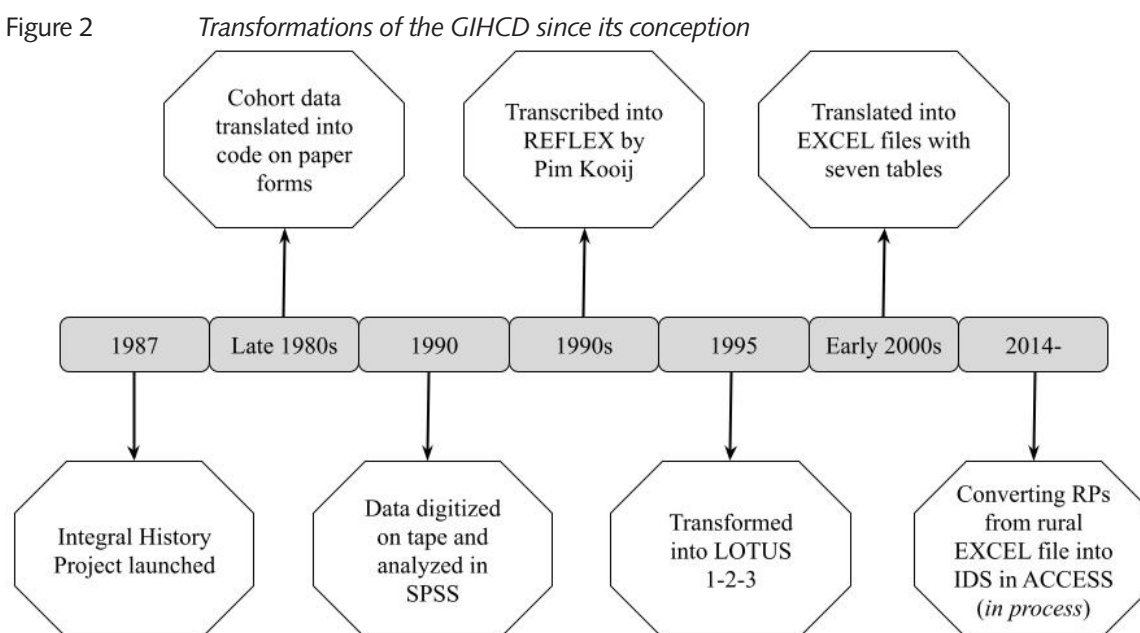
In the meantime, NWO funded in the years 1995–1996 a Dutch-Flemish research project concentrating on family strategies. Paping's part of the project used the Groningen Integral History Cohort Database (GIHCD) to study strategies regarding migration in the countryside. He had to conclude that — despite all the efforts — the cohort analysis database was still not very consistent and had significant gaps. The quality of the 1811 cohort was especially poor, partly due to the difficulties to detect the migration history of the cohort members, as the Dutch population registers keeping systematic track of migration only started in 1850.

As a preparation for the family strategy project, the original REFLEX-database was converted into a LOTUS 1-2-3 spreadsheet, making the data analysis and the input of extra data easier. However, one major weakness of the used program was that it had difficulty accepting dates before 1880. To solve this problem, 100 years were added to all the dates. In the end, dates in the LOTUS database were changed into relative parts of the year — for instance, the 30th April 1840 became 1840.33 — making it easy to calculate time spans. This implied that the exact dates got lost, although these are still available on the original forms. Next, all the data of all RPs from the nine rural municipalities and the cohorts 1830, 1850 and 1870 were checked by Paping, focussing on RPs with an incomplete life course. It turned out that previous investigators had missed a lot of information, which sometimes could even be found at the place of birth. For the still numerous missing persons, it was checked if they had gone to the city of Groningen or if they appeared on American migration lists. Also, many lost cohort members were found by using newly available indices and films of civil certificates and by systematically scrutinising the whole collection of published genealogies in the provincial Groningen Archive. In general, cohort members were followed through the whole province of Groningen at least until 1900, but preferably until after 1920.

Consequently, by 1999 the quality of the 1830, 1850 and 1870 rural cohorts had been greatly improved. Of the 1830 and 1850 cohort members only 8% and 10% were lost somewhere during their life in the province of Groningen. The situation was worse for the 1870 cohort where about a quarter had not (yet) been traced during their life span; however, only 4% disappeared before 1900 (Paping, 1999, p. 79). These first overviews of the cohorts showed a massive and rising emigration of the cohort members out of the province of Groningen, respectively 11% of the 1830 cohort members, 18% for 1850 and even 26% for 1870. These observed migration shares are even more impressive taking into account the high child and juvenile mortality — as RPs dying early obviously did not have much time to move out of the province — and the percentages of lost persons just mentioned.



Figure 2



In 2002, in line with these interesting migration patterns, the Faculty of Arts of the University of Groningen financed a research project following the migration all over the Netherlands. The research period was extended until 1940 and in first instance extra information on migrations within the province of Groningen was gathered. Subsequently, RPs were followed outside the province, by visiting the municipal archives of main destinations like Amsterdam and writing letters to the smaller archives. This research effort again greatly improved coverage and reliability of the GIHCD, which by then had been converted from LOTUS into EXCEL. However, the extremely time-consuming gathering of data on all migrated individuals, was one of the major causes for this project not to have resulted in an end publication.

Since the millennium change, digitisation of the civil registration has rapidly proceeded. Thanks to the efforts of the Groningen archives, indices with the names, ages and occupations mentioned in all birth, death and marriage records became increasingly available via the website [AlleGroningers](#) (see also [Mandemakers, Bloothoof, & Laan, forthcoming](#)). Numerous volunteers made the transcripts and improved the Groningen Archives database, adding extra years when allowed by Dutch privacy regulations (public after 100 years for births, 75 years for marriages and 50 years for deaths). A more recent development was the addition of the scans of the original sources. Since the early 2000s, a sophisticated online search engine made it possible to trace individual persons from behind the computer! It became much easier to establish if a certificate really was about the person one was looking for, and this without the time-consuming travel to any archive. One can also search for a combination of names which makes it easy to find specific couples. Also, for other parts of the Netherlands search engines became available to tackle huge databases — especially [WieWasWie](#) covering the whole of the Netherlands, though also search engines from separate archives — making it increasingly easier to track down lost cohort members.

In conclusion, digital tools have made it much easier from about 2005 onwards to develop the GIHCD. These digital improvements have been especially useful for improving the quality of the 1811 rural cohorts, which had been largely neglected since 1995, as well as the 1811, 1830, 1850 and 1870 cohorts of the city of Groningen. However, due to lacking research funds the Integral History Project was still unable to do a general update of all cohorts.

Several history students wrote their master or bachelor thesis using the GIHCD: Leendert Klokkenburg (2009) discussed differences between orthodox and non-orthodox Calvinist rural RPs, Bart Hoogenboom (2013) investigated migration from the Groningen countryside to the city of Groningen and Piet-Jan Koning (2019) researched the huge migration from the Groningen countryside to the United States of America (USA). Koning showed that it is possible to trace at least three quarter of the migrating rural RPs born in 1850 and 1870 in America using among others the digitalised US census and passenger list data available online. The relatively small sample size allowed Koning to search for these lost RPs by hand and ultimately led to the retracing of 152 of 219 RPs studied that moved to the US. This seems better than software-based approaches used in recent initiatives to match and validate almost 500 persons of the HSN who migrated to the USA using American censuses (see [Paiva, Anguita, & Mandemakers, 2020](#)). So, retracing lost RPs in the USA manually might be a fairly efficient project when conducted on a small scale.

Another stimulus for the GIHCD was its participation in the European Historical Population Sample Network (EHPS-Net) in the period 2011–2016. Jacek Pawlowski was hired in 2014–2015 to put the data of the 1830, 1850 and 1870 RPs in the Intermediate Data Structure (IDS; [Alter & Mandemakers, 2014](#)). In this period a general update of the database took place by digitally tracing part of the remaining lost cohort members using newly available search engines.

## 3 CONTENT AND STRUCTURE OF THE GIHCD

### 3.1 OVERVIEW

As has been explained in the previous section, a major part of the Integral History Project focused on sampling microlevel data for the province of Groningen. The GIHCD sampled data on RPs born between 1811 and 1870 based on birth certificates. However, the history, the scope, the data collection and the sampling strategy was much different from that performed by the Historical Sample of the Netherlands (HSN, see [Mandemakers, 2000](#)).

The life course data is mainly, but not exclusively, drawn from civil registration documents (birth, death and marriage records) that describe individuals' life events. For the period after 1850 — when the Dutch state required municipalities to systematically keep track of their population — population registers have also been used. Ultimately, of a total of 5,280 RPs involved, information on 3,240 of them, derived from the nine rural municipalities (see figure 1), has been converted into an IDS database ([Alter & Mandemakers, 2014](#)). The dynamic nature of the Dutch population registers allowed documenting a number of characteristics of the RPs — like place of residence and occupation — throughout their lives.

As the database does not contain identifiable personal information relating to RPs younger than 100 years and the data of other persons such as children of the RPs are only anonymously processed, there are no impediments regarding privacy. At the moment, researchers can only get the database on request, though it is the aim to make it downloadable in the near future starting with the parts converted into the IDS format. This would allow researchers to use the sample for various studies on demographic and socioeconomic history, as the database includes precise and complete information on the RPs' occupations and migration history as well as some information on their family members. In section 3.2, we will describe in detail the sample design and content of the database and in section 3.3 we focus on the IDS component, which is also the most complete part. When relevant we also reflect on those parts of the database that are still compiled in an EXCEL file. An earlier description of the GIHCD can be found online.<sup>3</sup> This description provides a detailed introduction to the data included in the database.

Besides the GIHCD, a related and partly supplementary database is also available. The Database Roman-Catholics Groningen consists of a family reconstitution of all Roman-Catholics — about 5% of the population — roughly living in the 'GIHCD-part' of the Groningen countryside in the 18th century ([Paping, 2009](#); [Paping & Collenteur, 2004](#); [Paping & Schansker, 2013](#)). This database — the first version of which was constructed for the master thesis of the first author of this article — offers systematic information on the life courses of about 5,000 persons born between 1721 and 1810.

### 3.2 SAMPLE DESIGN, SOURCES AND CONTENT

The GIHCD focuses on the large city of Groningen and on nine of the 56 rural municipalities that made up the province of Groningen in the 19th century. The nine municipalities are Zuidhorn, Hoogkerk, Leens, Uithuizen, Bedum, Appingedam, Stedum, Beerta and Winschoten. Together they comprise nine of 36 rural municipalities in the clay parts of the province. The selection method of the 10 research municipalities has already been discussed in section 2.1, where a map is also presented (figure 1). Table 2 shows some characteristics of these municipalities.

Using the 1862 agricultural statistics ([Bijdragen, 1870](#), see Table 2), the share of members of farm labourers' families in the total population of the clay parts of Groningen can be estimated as 41%, while its unweighted average share in the selected nine rural municipalities is only 38%. This figure is a little bit lower than the total average for the clay region, because we have included in our rural sample two relatively urbanised

3 See "Integral History Project Groningen", *EHPS-Net*: <https://ehps-net.eu/databases/integral-history-project-groningen>, assessed on 20 April, 2022.

municipalities, Appingedam and Winschoten. For members of farmers' families, the shares in total population are 19% and 20% respectively. The remaining 40% of the population consisted mainly of the families of artisans, millers, shopkeepers, innkeepers, merchants, shippers, reverends, schoolmasters and others active in industry and services. This high non-agricultural share denotes the large extent of specialisation of tasks within the regional economy, which was combined with a very market-oriented agriculture. Before the end of the 19th century, large-scale factories were rare in the clay parts of the province of Groningen (Paping, 1999).

The initial sampling of RPs was based on the birth certificates, being available from August 1811. The sample was drawn with intervals of about 20 years, which led to four cohorts of individuals born in 1811, 1830, 1850 and 1870, though some cohorts include some extra years, due to the small size of some of the municipalities involved. For every municipality, the first 120 registered births in each cohort were used. For the large city of Groningen 240 certificates were selected, using the first 20 of every month.

The selected RPs were subsequently traced along specific moments during their life course, such as their wedding, death and birth of their children. The sources which were systematically checked were the birth, marriage and death certificates and the population registers starting in 1850. Other sources have been used to corroborate existing information or for tracking a person who could not be found in the mentioned ones. Examples of these sources include the census lists of 1815, 1829/1830 and 1839/1840, genealogies, military draft records, migration lists and tax records.

Unfortunately, due to the lack of systematic household information on the forms (partly also due to the limited availability of this information in the population registers before 1862), the database does not contain data on the precise household composition over time. It only provides information on the parental or marital relationships between individuals. This allows researchers to analyse families, but not households. However, this might be less of a problem, as the dominant household form in the province of Groningen was the nuclear family, and estimates are available on the date when the RP left the parental home (Paping, 2004b, pp. 278–279; Paping, 2018).

For the cohorts that were drawn from the city of Groningen more linkages are missing (see Appendix A1–4), although recent efforts have improved them to a considerable extent. Tables 3 and 4 provide an overview of how far RPs could be followed during their life course. As the information regarding the urban cohorts has been digitalised only partly, these cohorts have been excluded from tables 3 and 4. The lower result for the 1811 cohort is partly caused by the relatively high number of inaccuracies in the 1811 birth registration, as these were the first official birth records in the new Dutch civil registration.

Table 2 *Description of the municipalities selected by the Integral History Project Groningen*

Municipality	Description	Population 31 Dec. 1829	Population 31 Dec. 1899	% farmers 1862	% farm labourers 1862
Groningen	Large urban centre	30,260	66,537	0%	0%
Appingedam	Small city & 5 hamlets	2,855	4,348	14%	15%
Bedum	Large village & 5 hamlets	2,720	4,894	29%	40%
Beerta	2 larger villages	2,777	4,123	15%	42%
Hoogkerk	3 villages	815	2,082	41%	30%
Leens	6 villages	2,736	3,902	16%	44%
Stedum	3 villages	1,354	2,257	20%	54%
Uithuizen	Large village	1,921	3,730	13%	49%
Winschoten	Urbanised village	3,229	9,668	9%	13%
Zuidhorn	2 larger villages	1,720	2,840	16%	44%

*Explanation: The percentage of the members of farm labourers' families (including farm servants) and of the farmers' families in 1862 (based on data from Bijdragen, 1870) is calculated as a share of the total population of a municipality according to the census of 31 December 1859 ([www.volkstellingen.nl](http://www.volkstellingen.nl)).*

Table 3 *Life course characteristics of the 1811–1870 'rural' cohorts (n = 1,080 per cohort)*

	1811	1830	1850	1870	Total
Lost	2.1%	1.8%	1.2%	0.6%	1.4%
Emigrated permanently	4.2%	8.0%	11.5%	16.7%	10.1%
Died before 20 in the Netherlands	28.0%	25.6%	29.0%	33.5%	29.0%
Died after 20 in the Netherlands	65.7%	64.6%	58.3%	49.2%	59.5%
Marriage year known	56.1%	58.4%	55.8%	54.2%	56.1%

Notes: The 1811 cohort has not been converted into IDS yet. For a data overview per municipality, see Appendix A.

Table 4 *State of the life courses in the rural part of the GIHCD, per age group for the total of the cohorts 1830, 1850 and 1870 (n = 3,240).*

	Migrated permanently	Died within the Netherlands	Lost in research process	Remaining in the database
Aged 0–20	4.1%	29.4%	0.6%	Age 20: 66.0%
Aged 20–40	5.7%	11.6%	0.5%	Age 40: 48.2%
Aged 40–60	2.0%	10.4%	0.1%	Age 60: 35.7%
Aged 60+	0.3%	35.4%	0.0%	–

Table 3 shows that only a few RPs were lost without a trace. Even the cohort of 1811 seems relatively good. So, for all cohorts most of the life course of the RPs is known. The reported huge death rate of juveniles in 1870 can be contributed to a smallpox epidemic, which substantially increased infant mortality, making the choice for this specific sample year rather unfortunate. Both the increasing child mortality and the increase in emigration — not all the marriages abroad have been tracked yet — resulted in a slight decline of the number of known marriages in the database between 1830 and 1870.

As explained in section 2.3, from 1995 onwards the research effort was primarily focused on the rural 1830, 1850 and 1870 cohorts. This effort concentrated on the manual linking of the RPs with their relatives (parents, spouse and children). By doing intensive research in online databases and other sources of information for missing information on parents, marriage, children death and migration, nearly all RPs have been successfully linked with their relatives. Table 3 shows that for the 1811–1870 cohorts, we have no death certificate or emigration data for on average 1.4% of the 4,320 RPs, ranging from 2.1% for the 1811 cohort to 0.6% for the 1870 one. For the 1830–1870 cohorts this is even only 1.2% (table 4). Including also parents, husbands and children the last cohorts contain information on 19,045 individuals.

Recently collected information on the fate of most of the RPs who emigrated to the USA has been added to the EXCEL database, but not yet to the IDS version (especially marriage dates, characteristics of partners and death dates). The USA was a popular destination among lower-income groups in the second half of the 19th century, especially during the agrarian depression in the 1880s and early 1890s. In the future, more information on the life courses abroad will be added to the database. This addition will allow for new avenues of research that have not been explored yet, such as the socioeconomic success or failure of RPs who moved abroad. Some first results were presented by Koning and Paping (2019) showing a relatively huge upward social mobility of these American migrants.

### 3.3 STRUCTURE OF THE DATASET

Three cohorts of the rural part of the database (1830, 1850 and 1870) have been converted into the so-called Intermediate Data Structure (IDS; see Alter & Mandemakers, 2014). The IDS has become an increasingly popular format for life course databases (Dribe & Quaranta, 2020; Edvinsson & Engberg, 2020; Jenkinson, Anguita, Paiva, Matsuo, & Matthijs, 2020). Databases in IDS format allow researchers to conduct international and interregional studies with more ease, as it standardises data across databases. The IDS is designed in six tables which key structure can be used in a database management system to connect individuals with each other and with the contexts they are part of on specific moments in time. Until now five out of these six tables have been used for the GIHCD in a Microsoft Access database: INDIVIDUAL,

INDIV\_INDIV, CONTEXT, CONTEXT\_CONTEXT and METADATA, excluding the table INDIV\_CONTEXT. The information included in the five tables is presented below.

The INDIVIDUAL table includes information on personal attributes (e.g., name, occupation) and events (e.g., birth date, marriage date). The basic structure of the table includes a database identifier (*Id\_D*) and an individual identifier (*Id\_I*) for all included individuals, in our case each RP and the direct relatives that we found and linked to the RPs. Table 5 shows the attributes and events recorded in the GIHCD.

The INDIVIDUAL table counts 90,592 records unevenly belonging to 19,045 unique persons (RPs, parents, spouses and children). Each record consists of a value for a specific type of attributes. Examples of attributes are the RP's first name, the last name and the location of birth. The *Type* Birth\_Location can take the *Value* 'Hoogkerk', one of the municipalities. The relatives on which the database provides information is limited to the RPs' parents, their eventual children and marriage partner(s). The information available on these relatives is limited but include sex, birth location, birth date and occupation (only for parents and partners). All occupations are provided with a code number from the Historical International Classification of Occupations (HISCO; van Leeuwen, Maas, & Miles, 2002). Occupational information was collected from civil certificates. For example, the occupations of the parents of an RP are collected from the birth certificate and the occupation of the RP's partner is taken from the marriage certificate. Because more databases also draw on the HISCO codes to make claims about the social position of individuals and their parents, the addition of these codes will allow researchers to easily use the IDS release of this database for comparative research (compare Edvinsson & Engberg, 2020; Mandemakers & Kok, 2020; Vézina & Bournival, 2020).

Finally, the INDIVIDUAL table provides a context identifier (*Value\_Id\_C*) to connect contextual data to the CONTEXT table. Table 5 outlines some of the aforementioned information that can be found in the INDIVIDUAL table, excluding the time stamps. Three marginal departures from the IDS guidelines stand out. Firstly, in case of the *Type* "Departure to" we have also filled in the name of the location to which RPs moved next to a context identifier. Secondly, we used the value "IntGron\_form" for the field *Source* (see Table 6). This value stands for the forms that were used in collecting the data. As mentioned in section 2.3, researchers frequently failed to write down all information on the RPs in the formative years of the database. Source specification was one of the fields that researchers sometimes left empty, which resulted in source specifications with the *Value* "IntGron\_form" (3,350 records out of 90,592). Thirdly, in the IDS format, we could not specify the source in case of the *Type* "End\_Observation". Contrary to "Start\_Observation", for which we automatically could fill in the birth certificate as the only possible source value, the source specification for the *Type* "End\_Observation" was sometimes problematic since there are different ends of the observable life course. Observations may end when someone passes away (*Value*: "Death"), when the RP departs out of the register (*Value*: "Departure") to a location in which she or he is not found again, or when the RP is no longer found in registers after being present at the closing of previous registers (*Value*: "End source"). So, for the *Type* "End\_Observation" the values for *Source* are empty. See Table 6 for examples of records in the table INDIVIDUAL.

One major departure from the IDS format is the way in which changes in occupation are sometimes denoted. When time stamps were unavailable, the GIHCD indicates changes in occupations by a new occupation *Type* ("Occupation1", "Occupation2", "Occupation3", etc.). Here, "Occupation1" denotes the RP's occupation before and at marriage and, wherever applicable, "Occupation2" and higher denote changes in occupations after marriage. This choice was made because several of the paper forms on the basis of which the database is constructed failed to report the dates on which a new occupation was assigned or reported after marriage. For some occupations this exists on the paper forms but the dates have not been digitised yet. Therefore, it was not possible to indicate the timing of occupation changes, but it was required to use the new developed occupation *Types*. This is something that needs to be adjusted in the future to ensure that the GIHCD data can be combined with data from other IDS databases.

The INDIV\_INDIV table records the relationship between the individuals in a database which was in our case restricted to the relations between the RPs and his/her parents, spouse(s) and children. This table uses a second individual identifier (*Id\_I\_2*) which links the second individual to the first individual (*Id\_I\_1*) in each row. The field *Relation* contains a value that describes the relationship between both individuals (Alter & Mandemakers, 2014). In Table 7 the structure of the relationships of the first RP of our database is presented, excluding time stamps.

Table 5 *Attributes and events covered in the INDIVIDUAL table*

Attribute or Event	Type name	For RPs	For RPs' Partner(s)	For Parents	For Children
First name	First_name	X			
Last name	Last_name	X			
Sex	Sex	X	X	X	
Date of birth	Birth_date	X	X	X	X
Place of birth	Birth_location	X	X	X	X
Marriage date(s)	Marriage_date	X	X	X	
Marriage location(s)	Marriage_location	X	X		
Date of death	Death_date	X			
Place of death	Death_Location	X			
Location of migration	Departure_to	X			
Occupational titles	Occupation, Occupation1, Occupation2, etc.	X	X	X*	
HISCO code	Occupation_HISCO, HISCO1, etc.	X	X	X	

\* *The only occupation of the parents is the one reported on the birth certificates of the RPs.*

Table 6 *Examples of information in the table INDIVIDUAL*

ID	Id_D	Id_I	Source	Type	Value	Value_Id_C
1	IntGron 2015_02	1	Birth certificate	Birth_Date	<time stamp>	
5	IntGron 2015_02	1	Birth certificate	Birth_Location	Hoogkerk	19
2	IntGron 2015_02	1	Death certificate	Death_Date	<time stamp>	
130961	IntGron 2015_02	1	Population register	Departure_To	Bedum	11
132973	IntGron 2015_02	1	Population register	Departure_To	Hoogkerk	19
29161	IntGron 2015_02	1	Death certificate	End_Observation	Death	
4	IntGron 2015_02	1	Birth certificate	First_Name	Goossen	
3	IntGron 2015_02	1	Birth certificate	Last_Name	Aalfs	
77772	IntGron 2015_02	1	Marriage certificate	Marriage_Date	<time stamp>	
82516	IntGron 2015_02	1	Marriage certificate	Marriage_Location	Bedum	11
106307	IntGron 2015_02	1	Marriage certificate	Occupation1	Zonder	
106308	IntGron 2015_02	1	IntGron_form	Occupation2	landbouwer	
6	IntGron 2015_02	1	Birth certificate	Sex	Male	
22681	IntGron 2015_02	1	Birth certificate	Start_Observation	Birth	

In the CONTEXT table descriptive information of the selected municipalities, the city of Groningen and all the localities appearing in the sources as locations of RPs residence, is stored. The attributes (*Type*) are limited to the names, longitudinal centroid, latitudinal centroid, and the type of locality (hamlet, village, town, city, or municipality) and a context identifier (*Id\_C*). The spatiotemporal data is drawn from the Dutch Toponyms Spatio-Temporal 1812–2012 database of the HSN (Huijsmans, 2013). Households have not been added, nor families; even though the data from the INDIV\_INDIV table allows for some basic family reconstitutions which could include the parents and children of RPs but no other kin, such as RPs' siblings.

The CONTEXT\_CONTEXT table embeds the villages, hamlets and towns in the municipalities they are part off. The designation of these localities to their corresponding municipality is also based on the Dutch Toponyms Spatio-Temporal 1812–2012 database. Similar to the INDIV\_INDIV table, this table links localities to their municipality by linking two identifiers (*Id\_C\_1* and *Id\_C\_2*) and defining the relationship (field *Relation*) between the context layers.

Table 7 *Example of records in the table INDIV\_INDIV*

ID	Id_I_1	Id_I_2	Id_D	Source	Relation
1823265	1	3241	IntGron 2015_02	Marriage certificate	Husband
1824949	3241	1	IntGron 2015_02	Marriage certificate	Wife
1794525	1	6308	IntGron 2015_02	Birth certificate	Father
1801495	6308	1	IntGron 2015_02	Birth certificate	Child
1796648	1	8866	IntGron 2015_02	Birth certificate	Father
1801496	8866	1	IntGron 2015_02	Birth certificate	Child
1815102	1	12563	IntGron 2015_02	Birth certificate	Child
1808622	12563	1	IntGron 2015_02	Birth certificate	Father
1818342	1	15803	IntGron 2015_02	Birth certificate	Child
1811862	15803	1	IntGron 2015_02	Birth certificate	Mother

Although the specifications for the IDS (Alter & Mandemakers, 2014) emphasise that the INDIV\_CONTEXT table should connect individuals to their temporal and spatial contexts, such as addresses or municipalities, this table has not been filled out so far. The inclusion of information regarding individuals' migrations has been limited to the INDIVIDUAL table by the *Type* "Departure\_to". We do not expect this to be an issue for researchers as the only contexts to which RPs can be linked to are limited to hamlets, villages, towns and municipalities and not to lower-level contexts such as households, addresses or families. This means that "Departure\_to" captures the full extent of an RPs contextual changes. However, as shown by Alter, Newton, and Oeppen (2020) there are ways to create a context out of a nuclear family that is compatible with IDS, which could also be done for this database. Adding an INDIV\_CONTEXT table that connects individuals to families could open up new research avenues in the future.

The METADATA table is the fourth version of the IDS (4.01) with the addition of six new occupation *Types* ("Occupation1"–"Occupation6") which are employed by the GIHCD. When time stamps for all available occupations are traced, these additional *Types* will be replaced by time stamped occupations.

Finally, time stamps follow the latest IDS guidelines. *Date\_Type* indicates whether the date is an event that is observed on the event itself, like a birth date on a birth certificate, whether it is reported on a later date, or declared at a point for which an attribute is valid or assigned. *Estimation* indicates whether a date is exact, a middling year between two possible dates, a year, or period with a *Start\_Year* and an *End\_Year*.

### 3.4 STRENGTHS AND WEAKNESSES OF THE GIHCD

Presently, for only 39 out of the 3,240 rural cohort members of 1830, 1850 and 1870 no death date or migration outside of the Netherlands has been found. This means that the fate of slightly more than 1% of them is still missing (see tables 3, 4 and the Appendix). The most recent tracings of these RPs suggest that presumably a large part of them will have left the country or died on sea. This makes the GIHCD of excellent quality, at least when it comes to the coverage of the life courses of the persons involved. The extensive research process in the last decades has shown that in particular those RPs with a diverging life course were difficult to follow, for instance those migrating over larger distances and/or remaining unmarried, born outside marriage and those changing names.

Weak points of the GIHCD are: 1. Part of the information is still not digitalised and available on paper only; 2. Due to the changes in the collecting method, some information has not been systematically recorded, as for instance the changes in the precise household structure of the cohort members; 3. Although there is a lot of new information collected on the urban cohorts (4 \* 240 = 960 RPs) and the 1811 rural cohorts (1,080 RPs), this has not yet been systematically included in the digital database; 4. The limited embedding in an organisational structure since the last 25 years, makes the database the sole responsibility of one person.

## 4 OUTPUT

As previously stated, the impact of the GIHCD has been limited, mainly because it has only been used by economic and social historians of the University of Groningen. Consequently, the studies that have been conducted so far refer to debates in historical demography and economic and social history. We will briefly address them in this section.

The first publication that made use of GIHCD data was the volume on the village of Hoogkerk between 1770 and 1914 (Kooij, 1993). Hoogkerk was selected for a pilot study in which various relevant questions and methods could be tested, especially because of its spectacular transformation from an agrarian village to an industrialised area around 1900. The book ambitiously aimed at integrating the various domains that explained social developments in the village through a lens that combined economic, political, social, cultural, religious and demographic elements. With the help of the GIHCD, the demographic and socioeconomic developments could be disentangled to some extent. Although the RP sample for the four cohorts from Hoogkerk was small ( $n = 480$ ), it was used by Marcel Clement (1993) to analyse demographic developments in Hoogkerk on a micro level. He studied (child) mortality, nuptiality, migration and social mobility. Due to high levels of child mortality and migration, only a small amount of the RPs delivered substantial data in Hoogkerk. But it showed how mobile 19th-century Dutch people were, as almost every RP left Hoogkerk at least once.

Within the field of historical demography, the interest in individual agency grew in the last decades of the 20th century. How and why families and individuals make choices or were forced to do so in order to improve their socioeconomic position (Engelen, Knotter, Kok, & Paping, 2004). This turn towards agency stimulated the analysis of short- and long-term decisions of households, families and persons, taking into account the wider historical context. However, quantitative data that only present static information on peoples' location or occupation, and do not show how these changed over time, is not suited to easily assess the motives behind the choices. As a possible solution, Paping (1999, pp. 18–19) used the GIHCD to explain how and why different socioeconomic classes made different short- and long-term choices regarding their employment, the employment of their children and their migration patterns. The study showed that married couples often migrated within the first years after marriage. Afterwards their inclination to migrate diminished rapidly, suggesting a rise in local social embedment over time. Furthermore, it was shown using the information on moments of leaving and juvenile occupations in the GIHCD that relatively many lower-income (unskilled) families opted for short-term strategies, like sending children away as live-in servants, which proved not very beneficial in the long run (Paping, 2004a, p. 188). For lower-income families, the employment of children as servants from the age of about 14 was one of the few options available to avoid the costs of having largely unemployed older children at home. Higher income groups, such as large farmers and the self-employed in industry and services, could afford to keep their children at home. For every social group, those children remaining at home proved to acquire on average a much better position in the long run than those becoming servants. That unskilled labourers responded to real wage increases at the end of the 19th century by increasingly keeping their children at home, suggests that they saw this choice indeed as a positive long-term strategy (see also Paping, 2017). However, changing views on education in the late 19th century might also have played a role (Kooij, 2004, p. 196).

In another publication, Paping (2004b) used part of the cohort database to examine the strategies families employed with regards to the labour of their family members. With the addition of financial microdata from various Nieuw Scheemda farmers' bookkeeping, Paping focused on the group of unskilled farm labourers in Groningen. The financial situation of unskilled worker families proved to be very volatile in the second half of the 19th century. With the exception of the male household heads, the rest of the family members — both wives and children — heavily depended on casual or seasonal labour. Adolescent sons and daughters usually left the household rather soon to become live-in servants. Information on the dates of marriage and the birth of the first child in the database showed that forced marriages due to pregnancies were common among unskilled worker couples. Such a situation restricted the opportunity for unskilled labourer families to make long-term decisions, as they would be forced to make short-term choices in response to an unplanned pregnancy.

About a decade after the pilot study on Hoogkerk, Kooij (2004) repeated some of the principles of the Integral History Project in a theoretical essay. One of these principles, the necessity to integrate the various domains that influenced peoples' lives had not changed. Kooij (2004, pp. 194) distinguished the economic, demographic, social, cultural, political and religious domains as central in understanding the pressures on individual and household decisions.



To analyse the relationship between two of these domains, the cultural/religious domain and the socioeconomic domain, Collenteur and Paping (2004) used rural cohort data from the GIHCD and tried to disentangle the connection between cultural and economic effects on peoples' decision to marry in a comparative contribution. They compared differences in marriage patterns between three Russian regions and two Dutch regions. The Russian regions consisted of the geographically farthest to the east, Tambov (located about 410 kilometers southwest of Moscow), the farthest to the west, Olonets (located about 180 kilometers northwest of St. Petersburg) and Yaroslavl (located about 250 kilometers northwest of Moscow). The Dutch regions were Groningen and North-Brabant. The data showed that, on average, rural Russians married at an earlier age compared to people from the Groningen urban and rural regions, but that differences between the Russian regions were huge. Furthermore, men and women in Groningen showed large variations in marriage age, as it was not uncommon to marry around the age of 20, although it was not rare for couples to wait until their 30s. Collenteur and Paping (2004) concluded that marriage patterns were likely driven largely by socioeconomic factors in Groningen, and RPs had a lot of agency with respect to their marriage age, at least in the absence of unplanned pregnancies. This was absolutely not the case in the Tambov region, the farthest east of the Russian regions studied. In Tambov, young adults usually married before their 20s with very small variation in ages at marriage, indicating an important role for cultural factors like traditions and norms influencing marriage decisions, leaving very limited room for individual agency. Although the differences between the Russian regions studied in this research proved to be large, the data suggest that the further east someone went, the lower the average age of marriage and the smaller the variation in age at marriage would be. In this way, Collenteur and Paping's research provide some support for Hajnal's (1983) hypothesis on marriage patterns and the so-called Hajnal line.

Kooij (2011) used the 1830 and 1870 cohorts of Beerta and Winschoten to analyse similarities and differences between a more urban-oriented region (Winschoten) and a completely rural region (Beerta) in the province of Groningen. He shortly compared various social groups on the basis of their life chances and mobility, and did not find major differences in life expectancy and geographic mobility between urban and rural environments. However, he concluded that for the 1870 cohort long distance-migration became more common, both within the Netherlands and outside of it. Furthermore, he showed that the agrarian depression of the late-19th century gave rise to chain migration towards the USA.

Finally, Paping and Pawlowski (2018) compared intergenerational occupational social mobility of rural-to-urban migrants with other groups, such as rural stayers, in the Groningen province. Their study used two databases. First, they employed the huge AlleGroningers database with summaries of all 234,000 marriages concluded in the province of Groningen between 1811 and 1934, out of which those of persons born in the Groningen clay area (see figure 1) were selected (N = 121,000). Second, the much smaller GIHCD was used, which allowed studying the relation between social mobility and rural-urban migration much more in depth, taking into account migration to more distant Dutch cities, as well as for instance migration taking place after marriage.

Rural males moving to an urban environment showed both much more upward and downward social mobility, compared to rural stayers or rural to rural migrants, who showed high social immobility. Furthermore, the findings reveal that urban *pull factors* were primarily responsible for the rural to urban migration, especially the better employment opportunities in the cities. This view contradicts a common explanation that rural-urban migration was mainly stimulated by bad rural circumstances forcing rural poor to move. Just the opposite took place: especially those descending from the somewhat higher social classes were much more inclined to move to the cities, as they had more useful skills than unskilled labourers for an urban environment. Although also experiencing relatively higher chances on downward mobility, migrants moving to cities showed, on average, even more upward social mobility than rural stayers in almost every social group. The GIHCD also made it possible to look at short stayers. According to some research (see Puschmann, 2015), positive results for rural to urban migration are biased upward because they do not account for migrants that return after a short stay in the city. In this view, a return to the rural environment likely represented a failed move to the city. Paping and Pawlowski (2018), however, show that even a short stay in an urban environment led on average to lasting positive effects on the social position for most socioeconomic groups.

## 5 CONCLUDING REMARKS

The success of the so-called cohort analysis in the Integral History Project was restricted by the combination of ambitions that were too high, some naïve decisions being made and the lack of structural funding,

in addition to the consequences of enormous and rapid technical changes taking place since the 1980s. Until 1995, the quality of the database was too low, as a consequence the research output was not very impressive. Later on, the quality improved and several meaningful scientific publications were based on it. However, the use of the database remained difficult due to the struggle to regularly update it technically, and to incorporate new digitally available sources, without proper funding.

Despite all these drawbacks, in one respect the quality of the GIHCD is extremely high. In contrast with other databases only less than 2% of persons involved could not be traced during their whole life course, making it very representative as it also includes the most extreme and rare cases that in other databases are often missing. This is a major reason to try to make the database better available in the future, both in an IDS structure and as an EXCEL spreadsheet.

The relatively small size remains problematic, especially in this time of digitalisation. It seems rather odd to study only a sample of the population, while — at least for part of the less complex questions — you can just as well use the enormous databases now available in the Netherlands. After all, as the quality of software-based record linkages is still improving, the future might not be for samples as the GHICD, but for enormous databases connecting events of all persons in a geographical region, at least for the 19th and 20th century (see for northwest Groningen: Paping & Schansker, 2013, 2014), and maybe even for an earlier period. However, as long as not all the relevant sources have been digitalised in a way that makes linking possible, and as long as digitalised linkage procedures are still not perfect, there is a need for high quality specific databases like the Groningen Integral History Cohort Database.

## REFERENCES

- Akolzina, M., Dyatchkov, V., Kanitshev, V., Kontchakov, R., Mizis, Y., & Morozova, E. (2004). A comparison of cohort analysis and other methods of demographic microanalysis used in studying the Tambov region. In P. Kooij & R. Paping (Eds.), *Where the twain meet again: New results of the Dutch-Russian project on regional development 1780–1917* (pp. 45–90). Groningen/Wageningen: NAHI. Retrieved from <https://ugp.rug.nl/ha/article/view/2103>
- Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal historical microdata, version 4. *Historical Life Course Studies*, 1, 1–26. doi: 10.51964/hlcs9290
- Alter, G., Newton, G., & Oeppen, J. (2020). Re-introducing the Cambridge Group family reconstitutions. *Historical Life Course Studies*, 9, 24–48. doi: 10.51964/hlcs9311
- Clement, M. (1993). Demografisch gedrag, leefsituatie en mobiliteit. Een analyse van vier generaties. In P. Kooij (Ed.), *Dorp naast een stad: Hoogkerk 1770–1914* (pp. 160–201). Assen: Van Gorcum.
- Collenteur, G., & Paping, R. (2004). Age at first marriage in eighteenth and nineteenth-century Russia and the Netherlands: Tradition or economic and social circumstances? In P. Kooij & R. Paping (Eds.), *Where the twain meet again: New results of the Dutch-Russian project on regional development 1780–1917* (pp. 147–167). Groningen/Wageningen: NAHI. Retrieved from <https://ugp.rug.nl/ha/article/view/2103>
- Commissie voor de statistieke beschrijving der provincie Groningen (Eds.). (1870). *Bijdragen tot de kennis van den tegenwoordigen staat der provincie Groningen* (Vijfde deel: Landbouw-statistiek). Groningen: Hoitsema.
- Dribe, M., & Quaranta, L. (2020). The Scanian Economic-Demographic Database (SEDD). *Historical Life Course Studies*, 9, 158–172. doi: 10.51964/hlcs9302
- Duijvendak, M. G. J., & Blijham G. J. (1994). Groninger provinciale financiën in Nederlands perspectief 1824–1910. *NEHA-Jaarboek voor Economische, Bedrijfs- en Techniekgeschiedenis*, 57, 206–248. Retrieved from <https://webstore.iisg.nl/neha/13805517-1994-001.pdf>
- Dyatchkov, V., Kanitshev, V., Mizis, Y., Orlova, V., Protasov, L., & Protasov, S. (1998). Cohort analysis of Malye Pupky's population: Some preliminary results. In P. Kooij (Ed.), *Where the twain meet. Dutch and Russian regional development in a comparative perspective 1800–1917* (pp. 141–154). Groningen: NAHI.
- Edvinsson, S., & Engberg, E. (2020). A database for the future. Major contributions from 47 years of database development and research at the Demographic Data Base. *Historical Life Course Studies*, 9, 173–196. doi: 10.51964/hlcs9305
- Engelen, T., Knotter, A., Kok, J., & Paping, R. (2004). Labor strategies of families: An introduction. *The History of the Family*, 9(2), 123–135. doi: 10.1016/j.hisfam.2004.01.001

- Golubeva, S. (1998). Age and patterns of marriage of Russian farmers in the Yaroslavl region. In P. Kooij (Ed.), *Where the twain meet. Dutch and Russian regional development in a comparative perspective 1800–1917* (pp. 169–174). Groningen: NAHI.
- Hajnal, J. (1983). Two kinds of pre-industrial household formation system. In R. Wall, J. Robin & P. Laslett (Eds.), *Family forms in historic Europe* (pp. 65–104). Cambridge: Cambridge University Press.
- Hoogeboom, B. (2013). *Van de Ommelanden naar de Martinistad. Migratie naar de stad als strategie om te ontsnappen aan relatieve rurale achteruitgang* (Unpublished bachelor thesis). University of Groningen, Groningen.
- Huijsmans, D. P. (2013). *IISG-LINKS Historische Nederlandse Toponiemen Sption-Temporeel 1812–2012. Release 2013.2*. Retrieved from <https://iisg.amsterdam/en/hsn/data/place-names>
- Integral History Project Groningen. (n.d.). Retrieved from <https://ehps-net.eu/databases/integral-history-project-groningen>
- 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*, 9, 197–217. doi: [10.51964/hlcs9301](https://doi.org/10.51964/hlcs9301)
- Klokkenburg, L. (2009). *Afgescheiden en afgezonderd? Het gedrag van de afgescheidenen in de Groninger klei gedurende de negentiende eeuw* (Unpublished master thesis). University of Groningen, Groningen.
- Koning, P. J. (2019). *Gouden bergen of de Groninger klei? Een case study over sociale mobiliteit van migranten uit de Groningse kleigebieden naar de Verenigde Staten tussen 1850 en 1940* (Unpublished master thesis). University of Groningen, Groningen.
- Koning, P. J., & Paping, R. (2019, November). Economic Refugees' from the northern Dutch countryside? Social origin and socio-economic success of Dutch migrants to the USA in the second half of the 19th century. Presentation at the *10th Day of the Historical Demography*, Nijmegen.
- Kooij, P. (1987). *Groningen 1870–1914. Sociale verandering en economische ontwikkeling in een regionaal centrum*. Assen/Maastricht: Van Gorcum.
- Kooij, P. (Ed.). (1993). *Dorp naast een stad: Hoogkerk 1770–1914*. Assen: Van Gorcum.
- Kooij, P. (1993a). Inleiding. In P. Kooij (Ed.), *Dorp naast een stad: Hoogkerk 1770–1914* (pp. 1–5). Assen: Van Gorcum.
- Kooij, P. (1993b). Bevolking: Huishoudens, gezinnen en sociale stratificatie. In P. Kooij (Ed.), *Dorp naast een stad: Hoogkerk 1770–1914* (pp. 130–159). Assen: Van Gorcum.
- Kooij, P. (Ed.). (1998). *Where the twain meet. Dutch and Russian regional development in a comparative perspective 1800–1917*. Groningen: NAHI.
- Kooij, P. (2004). Demographic development in the context of integral history, 1800–1917. In P. Kooij & R. Paping (Eds.), *Where the twain meet again: New results of the Dutch-Russian project on regional development 1780–1917* (pp. 191–198). Groningen/Wageningen: NAHI. Retrieved from <https://ugp.rug.nl/ha/article/view/2103>
- Kooij, P. (2011). Uit de klei getrokken. De levensloop van migranten uit de Oost-Groninger geboortecohorten Beerta en Winschoten 1830 en 1870. In T. Engelen, O. Boonstra & A. Janssens (Eds.), *Levenslopen in transformatie: Liber Amicorum bij het afscheid van prof. dr. Paul M. M. Klep* (pp. 279–290). Nijmegen: Valkhof Pers.
- Kooij, P., & Paping, R. (Eds.). (2004). *Where the twain meet again: New results of the Dutch-Russian project on regional development 1780–1917*. Groningen/Wageningen: NAHI. Retrieved from <https://ugp.rug.nl/ha/article/view/2103>
- Kooij, P., & Sleebe, V. (1991). A small village in a changing world; Integral history at a local level. *Economic and Social History in the Netherlands*, 3, 19–36.
- Mandemakers, K. (2000). The Netherlands. Historical Sample of the Netherlands. In: P. Kelly Hall, R. McCaa & G. Thorvaldsen (Eds.), *Handbook of International Historical Microdata for Population Research* (pp. 149–177). Minneapolis: Minnesota Population Center.
- Mandemakers, K., Bloothoof, G., & Laan, F. (forthcoming). LINKS. The LINKing System for historical family reconstruction in the Netherlands. *Historical Life Course Studies*.
- Mandemakers, K., & Kok, J. (2020). Dutch Lives. The Historical Sample of the Netherlands (1987–): Development and Research. *Historical Life Course Studies*, 9, 69–113. doi: [10.51964/hlcs9298](https://doi.org/10.51964/hlcs9298)
- Paiva, D., Anguita, F., & Mandemakers, K. (2020). Linking the Historical Sample of the Netherlands with the USA censuses, 1850–1940. *Historical Life Course Studies*, 9, 1–23. doi: [10.51964/hlcs9312](https://doi.org/10.51964/hlcs9312)
- Paping, R. (1995). *Voor een handvol stuivers. Werken, verdienen en besteden: de levensstandaard van boeren, arbeiders en middenstanders op de Groninger klei, 1770–1860* (Doctoral dissertation). Groningen: NAHI. Retrieved from <https://pure.rug.nl/ws/portalfiles/portal/14976871/thesispaping1995.pdf>

- Paping, R. (1999). Gezinnen en cohorten: Arbeidsstrategieën in een marktgerichte agrarische economie: De Groningse kleigebieden 1830–1920. In J. Kok, A. Knotter, R. Paping & E. Vanhaute (Eds.), *Levensloop en levenslot. Arbeidsstrategieën van gezinnen in de negentiende en twintigste eeuw* (pp. 17–88). Groningen/Wageningen: NAHI. Retrieved from [https://www.researchgate.net/profile/Jan-Kok-4/publication/333401946\\_Levensloop\\_en\\_levenslot/links/5cebd797a6fdcce250aa63ca/Levensloop-en-levenslot.pdf](https://www.researchgate.net/profile/Jan-Kok-4/publication/333401946_Levensloop_en_levenslot/links/5cebd797a6fdcce250aa63ca/Levensloop-en-levenslot.pdf)
- Paping, R. (2004a). Family strategies concerning migration and occupations of children in a market-oriented agricultural economy. *The History of the Family*, 9(2), 159–191. doi: 10.1016/j.hisfam.2004.01.003
- Paping, R. (2004b). Family strategies, wage labour and the family life cycle in the Groningen countryside, c. 1850–1910. In P. Kooij & R. Paping (Eds.), *Where the twain meet again: New results of the Dutch-Russian project on regional development 1780–1917* (pp. 271–291). Groningen/Wageningen: NAHI. Retrieved from <https://ugp.rug.nl/ha/article/view/2103>
- Paping, R. (2009). Gender and the intergenerational transfer of property and social position in the 18th and early 19th century northern Dutch countryside. In M. Durães, A. Fauve-Chamoux, L. Ferrer & J. Kok (Eds.), *The transmission of well-being: Gendered marriage strategies and inheritance systems in Europe (17th–20th centuries)* (pp. 291–313). Bern: Peter Lang. Retrieved from [https://www.researchgate.net/publication/260081918\\_Gender\\_and\\_the\\_intergenerational\\_transfer\\_of\\_property\\_and\\_social\\_position\\_in\\_the\\_18th\\_and\\_early\\_19th\\_century\\_northern\\_Dutch\\_countryside](https://www.researchgate.net/publication/260081918_Gender_and_the_intergenerational_transfer_of_property_and_social_position_in_the_18th_and_early_19th_century_northern_Dutch_countryside)
- Paping, R. (2010). Taxes, property size, occupations and social structure: The case of the 18th and 19th century northern Dutch countryside. *Belgisch Tijdschrift voor Nieuwste Geschiedenis*, 40(1), 215–248. Retrieved from <https://www.journalbelgianhistory.be/nl/journal/belgisch-tijdschrift-voor-nieuwste-geschiedenis-2010-1-2/taxes-property-size-occupations>
- Paping, R. (2017). Dutch live-in farm servants in the long 19th century: The decline of the life-cycle service system for the rural lower class. In J. Whittle (Ed.), *Servants in rural Europe. 1400–1900* (pp. 203–226). Woodbridge: Boydell & Brewer.
- Paping, R. (2018). The 'dynamics' of household structure and size in the Northern Dutch countryside around 1840. In P. Puschmann & T. Riswick (Eds.), *Building bridges. Scholars, history and historical demography. A Festschrift in honor of professor Theo Engelen* (pp. 362–384). Nijmegen: Valkhof Pers. Retrieved from [https://www.ru.nl/publish/pages/800120/building\\_bridges\\_open\\_access\\_pdf.pdf](https://www.ru.nl/publish/pages/800120/building_bridges_open_access_pdf.pdf)
- Paping, R., & Collenteur, G. (1998). The economic development of the clay soil area of Groningen 1770–1910: Income and socio-economic groups. In: P. Kooij (Ed.), *Where the twain meet. Dutch and Russian regional development in a comparative perspective 1800–1917* (pp. 35–50). Groningen: NAHI.
- Paping, R., & Collenteur, G. (2004). Population growth and social structure in a market-oriented agricultural economy in The Netherlands 1750–1820. *Obradoiro de Historia Moderna*, 13, 75–99. Retrieved from [http://dspace.usc.es/bitstream/10347/3936/1/pg\\_076-101\\_obradoiro13.pdf](http://dspace.usc.es/bitstream/10347/3936/1/pg_076-101_obradoiro13.pdf)
- Paping, R., & Pawlowski, J. (2018). Success or failure in the city? Social mobility and rural-urban migration in nineteenth- and early-twentieth-century Groningen, the Netherlands. *Historical Life Course Studies*, 6, 69–94. doi: 10.51964/hlcs9329
- Paping, R., & Schansker, G. (2013). The reproduction of the rural labour class: Fertility, nuptiality and life chances. Paper presented at the 6th Day of the Historical Demography. Retrieved from <https://www.researchgate.net/publication/282249917>
- Paping, R., & Schansker, G. (2014). De reproductie van de rurale arbeidersklasse in achttiende- en negentiende-eeuws Groningen: Vruchtbaarheid, nuptialiteit en overlevingskansen. In I. Devos, K. Matthijs & B. Van de Putte (Eds.), *Kwetsbare groepen in/en historische demografie* (pp. 71–98). Leuven/Den Haag: Acco.
- Puschmann, P. (2015). *Social inclusion and exclusion of urban in-migrants in northwestern European port cities. Antwerp, Rotterdam & Stockholm, ca. 1850–1930* (Doctoral dissertation). Retrieved from [https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1871846&context=L&vid=Lirias&search\\_scope=Lirias&tab=default\\_tab&lang=en\\_US&fromSitemap=1](https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1871846&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1)
- Schrover, M. (1998). Demographic behaviour in North Brabant in the nineteenth century. In P. Kooij (Ed.), *Where the twain meet. Dutch and Russian regional development in a comparative perspective 1800–1917* (pp. 201–222). Groningen: NAHI.
- Shustrova, I., & Sinitsyna, E. (2004). Demographic behaviour in the Yaroslavl loamy area. The results of cohort analysis for two typical rural villages. In P. Kooij & R. Paping (Eds.), *Where the twain meet again: New results of the Dutch-Russian project on regional development 1780–1917* (pp. 7–18). Groningen/Wageningen: NAHI. Retrieved from <https://ugp.rug.nl/ha/article/view/2103>
- Sinitsyna, E. (1998). Some results of cohort analysis in the Yaroslavl region. In P. Kooij (Ed.), *Where the twain meet. Dutch and Russian regional development in a comparative perspective 1800–1917* (pp. 155–168). Groningen: NAHI.

- Sleebe, V. (1993). Openbaar en privé. In P. Kooij (Ed.), *Dorp naast een stad: Hoogkerk 1770–1914* (pp. 295–335). Assen: Van Gorcum.
- Trienekens, G. (1987). Theoretische en methodologische aspecten van de lokale en regionale geschiedschrijving. In F. van Besouw, P. den Boer, F. W. N. Hugenholtz & Th. van Tijn (Eds.), *Balans en perspectief. Visies op de geschiedwetenschap in Nederland* (pp. 167–188). Groningen: Wolters-Noordhoff/Forsten.
- Trienekens, G. (1993). Integrale geschiedenis in wording: Aarle-Rixtel en Wanroij in de negentiende en het begin van de twintigste eeuw. In J. A. van Oudheusden & G. Trienekens (Eds.), *Een pront wijf, een mager paard en een zoon op het seminarie: Aanzetten tot een integrale geschiedenis van oostelijk Noord-Brabant* (pp. 211–313). 's-Hertogenbosch: Stichting Brabantse Regionale Geschiedoefening.
- Trienekens, G. (2004). Characteristics of households in the eastern part of North Brabant 1810–1920. In P. Kooij & R. Paping (Eds.), *Where the twain meet again: New results of the Dutch-Russian project on regional development 1780–1917* (pp. 25–44). Groningen/Wageningen: NAHI. Retrieved from <https://ugp.rug.nl/ha/article/view/2103>
- van Leeuwen, M. H. D., Maas, I., & Miles, A. (2002). *HISCO. Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.
- van Oudheusden, J. A., & Trienekens, G. (Eds.) (1993). *Een pront wijf, een mager paard en een zoon op het seminarie: aanzetten tot een integrale geschiedenis van oostelijk Noord-Brabant*. 's-Hertogenbosch: Stichting Brabantse Regionale Geschiedoefening.
- Vézina, H., & Bournival, J. S. (2020). An overview of the BALSAC population database. Past developments, current state and future prospects. *Historical Life Course Studies*, 9, 114–129. doi: [10.51964/hlcs9299](https://doi.org/10.51964/hlcs9299)
- Voerman, J. F. (2001). *Verstedelijking en migratie in het Oost-Groningse veengebied 1800–1940* (Doctoral dissertation). Assen: Van Gorcum.

## APPENDIX

Table A1 *Life course characteristics of the 1811 cohort*

1811	Total	Lost	Emigrated permanently	Died before 20 in the Netherlands	Died after 20 in the Netherlands	Marriage year known
City of Groningen	240	16	9	86	129	110
Appingedam	120	5	2	39	74	61
Beerta	120	2	5	37	76	66
Bedum	120	0	3	36	81	66
Hoogkerk	120	2	2	38	78	60
Leens	120	3	12	33	72	64
Stedum	120	4	6	34	76	66
Uithuizen	120	5	8	32	75	73
Winschoten	120	1	4	26	89	76
Zuidhorn	120	1	3	27	89	74
Total rural	1,080	23	45	302	710	606

Table A2 *Life course characteristics of the 1830 cohort*

1830	Total	Lost	Emigrated permanently	Died before 20 in the Netherlands	Died after 20 in the Netherlands	Marriage year known
City of Groningen	240	9	4	94	133	97
Appingedam	120	4	5	35	76	64
Beerta	120	3	4	20	93	83
Bedum	120	1	8	33	78	70
Hoogkerk	120	1	3	28	88	74
Leens	120	4	13	42	61	58
Stedum	120	1	15	33	71	72
Uithuizen	120	0	28	27	65	69
Winschoten	120	4	4	30	82	67
Zuidhorn	120	1	6	29	84	74
Total rural	1,080	19	86	277	698	631

Table A3 *Life course characteristics of the 1850 cohort*

1850	Total	Lost	Emigrated permanently	Died before 20 in the Netherlands	Died after 20 in the Netherlands	Marriage year known
City of Groningen	240	6	0	90	139	116
Appingedam	120	2	5	40	73	65
Beerta	120	1	7	30	82	71
Bedum	120	1	3	30	86	75
Hoogkerk	120	1	10	36	73	61
Leens	120	1	24	34	61	63
Stedum	120	1	21	43	55	62
Uithuizen	120	1	31	36	52	59
Winschoten	120	4	7	34	75	67
Zuidhorn	120	1	16	30	73	80
Total rural	1,080	13	124	313	630	603

Table A4 *Life course characteristics of the 1870 cohort*

<b>1870</b>	<b>Total</b>	<b>Lost</b>	<b>Emigrated permanently</b>	<b>Died before 20 in the Netherlands</b>	<b>Died after 20 in the Netherlands</b>	<b>Marriage year known</b>
City of Groningen	240	0	12	99	129	116
Appingedam	120	0	11	45	64	61
Beerta	120	2	25	40	53	67
Bedum	120	0	16	43	61	65
Hoogkerk	120	1	7	46	66	59
Leens	120	1	21	37	61	67
Stedum	120	0	39	44	37	61
Uithuizen	120	2	42	31	45	71
Winschoten	120	1	7	39	73	67
Zuidhorn	120	0	12	37	71	67
Total rural	1,080	7	180	362	531	585