

# Norwegian Historical Population Register, 1801–present

By Lars Holden, Svetlana Boudko and Bjørn Fjellvoll

To cite this article: Holden, L., Boudko, S., & Fjellvoll, B. (2026). Norwegian Historical Population Register, 1801–present. *Historical Life Course Studies*, 16, 58–73. <https://doi.org/10.52024/hlcs25461>

**HISTORICAL LIFE COURSE STUDIES**

VOLUME 16

2026



## MISSION STATEMENT

# HISTORICAL LIFE COURSE STUDIES

*Historical Life Course Studies* was established within *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.

### 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, the International Institute of Social History, the European Society of Historical Demography, Radboud University Press, Lund University and HiDO Scientific Research Network Historical Demography. 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:

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

&

Paul Puschmann (Radboud University)

### Associate Editors:

Gabriel Brea-Martinez (Open University of Catalonia & Centre for Economic Demography, Lund University)

&

Wieke Metzlar (Radboud University)



# Norwegian Historical Population Register, 1801–present

Lars Holden

Norwegian Computing Center

Svetlana Boudko

Norwegian Computing Center

Bjørn Fjellvoll

Norwegian Computing Center

## ABSTRACT

The Norwegian Historical Population Register (NHPR) reconstructs the entire Norwegian population from 1801 to the present and links it with the modern National Population Register (NPR). It provides a multigenerational research infrastructure spanning seven generations and functions as an authoritative identifier system for documenting individuals and connecting diverse archival materials. This article presents the structure and development of the NHPR's two components: the open online register (NHPR-O), which documents deceased individuals using publicly accessible sources, and the closed register (NHPR-C), which integrates historical and modern microdata under privacy-preserving conditions. We describe the core sources and the hybrid linkage system combining large-scale algorithmic matching with extensive crowdsourcing. The paper reports current linkage rates across censuses, discusses the completeness and quality of historical sources, and examines challenges such as duplicate registrations, and the difficulty of identifying individuals with sparse information. We outline how continuous crowdsourcing and iterative quality control steadily improve linkage accuracy. Finally, the article discusses representativity, the constraints of privacy legislation, and the potential for international cooperation and distributed population registers. Together, these developments establish the NHPR as a scalable and evolving resource for research, genealogy, and public use.

**Keywords:** Population register, Microdata, Census

e-ISSN: 2352-6343

DOI article: <https://doi.org/10.52024/hlcs25461>

© 2026, Holden, Boudko, Fjellvoll

This open-access work is licensed under a Creative Commons Attribution 4.0 International License, 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 Norwegian Historical Population Register (NHPR) covers the entire Norwegian population from 1801 to the present. It consists of a public register of deceased individuals documented in open sources, available at [histreg](https://histreg.no),<sup>1</sup> and a closed register that extends the National Population Register (NPR) and its identification system, which is used daily by all Norwegians. As a result, the NHPR can be combined with modern microdata. The main purpose of the NHPR is to serve as research infrastructure across disciplines, although it is also widely used in genealogical work.

Alter et al. (2023) and Mandemakers (2025) provide overviews of various longitudinal population databases, and the NHPR follows many of the same principles as similar databases (e.g. Mandemakers & Dillon, 2004). It is true to the source, complete, open, coherent, consistent, efficient and research-oriented (Vikstrøm et al., 2023). The NHPR also has several less common characteristics:

1. Covers the entire country
2. Extends the modern NPR
3. Can be combined with modern microdata with data minimisation
4. Is an authoritative register over individuals publicly accessible online
5. Incorporates as many sources as possible
6. Uses both crowdsourcing and algorithmic linkage
7. Is continuously improving and expanding

Full national coverage increases the feasibility of identifying individuals who migrate within Norway. Emigration protocols make it possible to track emigrants, while immigration to Norway are typically registered only when individuals appear in church records or censuses. This has been previously studied in Holden and Boudko (2018).

The NHPR serves as an authoritative register of deceased individuals, assigning a unique identifier (ID) to each individual. This ID is intended to serve as a standard identifier across all relevant types of texts, presentations and datasets and used as an identifier in as many archive sources as possible.

Primary sources prior to 1920 include censuses, church records, and emigration protocols, with additional primary sources becoming available after 1906. The NHPR also assigns unique identifiers (IDs) to individuals recorded in other sources related to education, health, incarceration, occupation, and politics. The aim is to gather as much information as possible about each individual, thereby increasing research value and enhancing the utility of all linked sources.

The NHPR project started in 2010 and the website [histreg](https://histreg.no) has been public since 2016. There has been a continuous development during all these years in development of the website, recruitment for crowdsourcing and algorithmic linking. Substantial effort was devoted to supporting the transcription of primary sources at the start of the project. Some of the sources, including the closed 1930 and 1950 censuses, are being transcribed using machine learning.

To understand the scale of the NHPR, a brief description of Norway's demographics is needed. The Norwegian population numbered approximately 0.88 million in 1801, rising to 2.2 million in 1900, and 2.6 million in 1920. Today the population stands at 5.5 million. The first census in Norway dates from 1801 and the second in 1865. The capital, Oslo, had 250,000 inhabitants in 1900 and today 725,000, within an expanded metropolitan area. Norway and Ireland had the highest relative emigration rates in Europe. Between 1820 and 1920, about 0.8 million Norwegians emigrated — primarily to North America — of whom roughly one quarter later returned to Norway, accompanied by approximately 9,000 children born abroad. The NHPR is documented in several papers including Thorvaldsen and Holden (2023), and Holden et al. (2025). The Nordic countries are known for high-quality register data with a homogeneous population.

The next section describes the sources and the different parts of NHPR. The linking methods are described in Section 3 followed by sections describing the open and closed parts of NHPR. The main objective of NHPR is as a research infrastructure. This is described in a separate section followed by a section on possible international collaborations.

1 <https://histreg.no>

## 1 SOURCES AND THE DIFFERENT PARTS OF NHPR

All the primary sources: national censuses, church books (before 1920), emigration protocols, and two paper-based registrations 1906–1995 and 1904–1960 respectively, are transcribed. However, National Archive of Norway has not published lists of the limited number of church books that are missing. All primary sources and most other sources are available in the Digital Archive ([Digitalarkivet.no](https://digitalarkivet.no)), which is managed by the National Archives. The Digital Archive assigns each individual named in a source a unique and permanent identifier, denoted PSID. Both the Digital Archive and NHPR use these PSIDs as part of the URL. On average, each individual is mentioned in about 10 sources in the period with open sources and this number is increasing after 1920.

New transcribed sources are continuously added, including thematic registers in the Digital Archive, such as probate, prison, health, and school records, as well as moving lists and the treason archive from the Second World War. Thematic registers that are not part of the Digital Archive, such as the Digital Archive of Norwegian Prisoners (1940–1945), the War Sailor Register, a register of politicians from Norway's independence in 1814 to the present, public servants 1867–1960, and material extracted from newspapers and biographical sources are also included. All individuals mentioned in these sources are also assigned PSIDs and may be linked.

We aim to include more archives into NHPR through collaboration with the owners of different archives (like the three first mentioned above) by enhancing the value of their collections and directing more visitors to their websites as part of our shared effort to make the NHPR as comprehensive as possible. NHPR is a register of individual records derived from archives and sources and is not intended to replace the different archives.

We aim to maintain a population register that is as open as possible, yet as closed as necessary to comply with privacy and legal regulations. Open data also enables the use of crowdsourcing and citizen science in development and quality assurance of the register. The sources may be divided into three main categories of openness:

1. OPEN: fully open and eligible for publication on the internet
2. RESTRICTED: researchers may obtain full access to the sources
3. CLOSED: researchers may in most cases only get access pseudonymised data, subject to data minimisation requirements

Closed data include censuses the last 100 years and church books: baptism last 100 years, confirmation last 86 years, marriages last 82 years and burial are open.<sup>2</sup> Further, the closed data include the entire NPR, most data regarding individuals who are still alive, some health data independent of the time of the source, a paper-based population register (1906–1995, nationwide from 1946), and a paper-based registration of births, marriages and deaths for the period 1904–1960. Several of these sources contain birth dates for several family members, which are invaluable for linkage. Before the project started, most of these paper-based closed sources were not transcribed.

An example for a restricted source is the treason archive from the Second World War. However, the names in this source are open. The international population data from IPUMS are restricted ([Ruggles et al., 2025](#)). The last large open sources are the 1920 census and death records until 2014. Access to open data is much less bureaucratic and easier to use in many research projects. It is possible to apply for access to data from the National Archive of Norway or Statistics Norway, but this process may be easier with a Norwegian partner. A Norwegian partner may also be important for having knowledge about and interpreting various sources.

Due to privacy restrictions we need to split the NHPR into an open and closed part: NHPR-O and NHPR-C. All records with deceased persons named in open sources are included in NHPR-O while all information is included in NHPR-C. Privacy regulations also require distributed storage for NHPR-C, which is not suitable for NHPR-O. Table 1 and Figure 1 give an overview of the sources and the three registers. Note that NHPR-O only covers the entire population until 1920 since the only large open source after the 1920 census is the death records.

2 <https://www.digitalarkivet.no/content/418/sperringsregler-ved-registrerte-historiske-dokumenter> (only in Norwegian)

Table 1 *The main differences between the three populations registers*

	NHPR-O	NHPR-C	NPR
Access	Open	Closed	Closed
Selection	Deceased, 1801–2014	All, 1801–present	All, 1964–present
Entire population	1801–1920	1801–present	1964–present
Size	1801–1920: 6.6 mill.	1801–1963: 9.1 mill.	Pop. 2026: 5.5 mill.
Source entries	>70 mill. only open sources	>200 mill.	Continuously updated
Data structure	Links to PSIDs in open sources	One of the IDs birthdate ID, NHPR-C-C or NHPR-C-O is added to each person record in the source	The birthdate ID is added to each person record in the source
Status	Available, continuously improved	First release summer 2027	Used since 1964

Note: The new IDs are explained in Section 5.

### 3 LINKING METHODS

Most links, e.g. identifying the same individual across different sources, are created by algorithms. More than 90% of the links in NHPR-O links are generated this way. The remaining links are made through crowdsourcing. This combination reflects the practical challenges of the linking process Holden and Boudko (2018).

Large volumes of the links are straightforward, for example the majority of the 2.1 million individuals registered in the 1910 and 1920 censuses, who have recorded birthdates and often appear in the same household. At the same time, there are many links that can only be established manually.

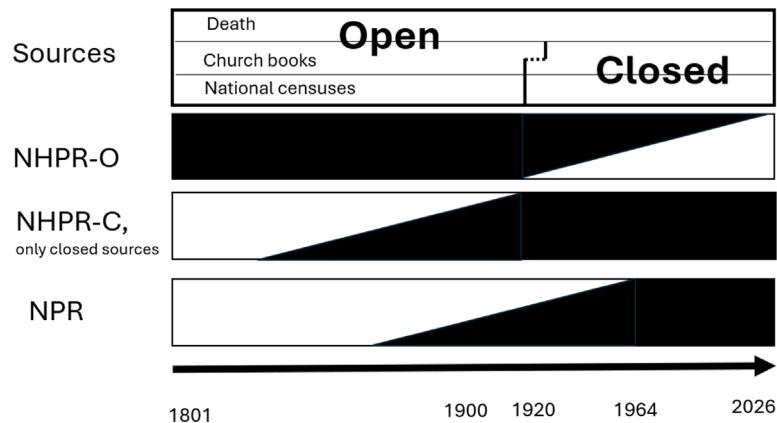
#### 3.1 CROWDSOURCING

More than 280 individuals have contributed to crowdsourcing by registering links since 2016 at the website [histreg.no](https://histreg.no). Contributors have been recruited since 2016 through presentations at local historical and genealogical societies, popular and scientific papers, exhibition, and media coverage. Some contributors are professionals working at archives, museums, and universities. There is no payment. Dialog on National Archives' discussion forum since March 2024 have been important for informing contributors and answering questions. Contributors may propose improvements and praise or criticise the system. In addition, there is almost daily email correspondence with contributors. These contributors are highly experienced and possess extensive knowledge of genealogy and the relevant sources. The crowdsourcing implies that the database is continuously growing and improving. The ongoing dialogue with contributors about the functionality of histreg, together with their suggestions for new features, effectively serves as a form of crowdtesting for the website. Because contributors continuously use histreg in various real-world scenarios, their feedback reveals bugs, usability challenges, and improvement opportunities that traditional in-house testing often fails to capture.

All links are registered with the contributor and a timestamp. The timestamps show that activity occurs across nearly all hours of the year. The process is somewhat like putting together a puzzle on the internet — it can be easy to become absorbed. Contributors often start by linking their own family and local community. Most manual links are created by 20–50 contributors, often retired individuals working several hours per day.

The histreg site also provides lists of person records that remain unlinked or where the system has flagged questionable links from a census in a municipality chosen by the contributor. This functionality enhances both representativeness and data quality while increasing contributors' efficiency. It enables contributors to focus on records that require manual inspection in municipalities and time periods where they have particular interest or expertise.

Figure 1 Overview of the sources and the three registers



Notes: The upper box shows the periods when these primary sources are closed. The three lower boxes show how large part of the population that is included in the three registers. We show NHPR-C limited to closed sources since NHPR-C covers the total population in the entire period. NHPR-C limited to closed sources covers the total population in 1920–present.

Crowdsourcing comes at the cost of not being able to enforce strict linking criteria such as identical birthdates, names, and regions. For this reason, linking must be iterative. Erroneous links are removed during regular quality controls, in contrast to many other population registers where links, once added, are almost never removed. If links are rarely removed, stricter rules must be applied for each new link, making it more difficult to achieve high linkage ratios.

While no formal statistical evaluation has been conducted, our experience indicates that manual contributors generally achieve lower error rates than algorithms. In January 2025 we sent 20 emails to contributors containing lists of 20–150 links made by each contributor that had been flagged as questionable by an algorithm. We received only positive responses to these emails, which might otherwise have been perceived as criticism. Most responses included detailed explanations of each link, and the contributor was correct in 20% of the cases. We concluded that the major contributors work very thoroughly.

Manual judgement is necessary when linking records with sparse information or unusual variations. For example, a difference of exactly 100 years in a birthdate probably indicates the same individual and an error in the source or transcription, while a difference of 7 years, 2 months, and 16 days most likely indicates different individuals.

A small proportion of links are manually marked as "certain" or "uncertain," depending on the data quality. "Certain" is applied to correct links where there are substantial deviations in characteristics such as name and birth year, to prevent manual or algorithmic delinking. "Uncertain" is used when the data are too sparse to establish a link at the expected level of quality, but where link retrieval may later be difficult. This may apply, for example, to an emigration record for a person known to have emigrated, where it is unclear whether a particular record in the emigration list corresponds to the correct individual. In January 2026, a warning was introduced on individual profiling pages at histreg where an algorithm questions the linking. The warning encouraged contributors to check the linking and either confirm it or remove erroneous links.

Family relationships are inferred either from source events (e.g. baptisms or household listings in censuses) or manually by contributors. We need to include an option for manually ignoring family relationships derived from church books, since in some cases stepparents may be registered as parents. This option is already included for censuses.

## 3.2 ALGORITHMIC LINKAGE

Several different algorithms are used to find links between person records in the open sources. Most algorithms attempt to link each individual in a source sequentially from the first person record according to certain criteria. The algorithms identify individuals and families with similar characteristics (names, birthdate/year, municipality, etc.) in different sources, either systematically between specific sources or

through an open search across all sources in the Digital Archive. The algorithms also consider existing links associated with the individual before proposing new links.

Quality-control algorithms systematically review all linked individuals and identify large deviations in characteristics, multiple registrations in the same source (such as a census, birth, or death record), or unlikely life courses. These algorithms automatically remove erroneous links or flag linked individuals for manual inspection. This creates a structured and dynamic interaction between automated processes and manual contributors.

Links in NHPR-O change continuously through the combined work of many manual contributors and several algorithms operating in parallel, each based on the current status of the database. It is therefore not feasible to shift to an improved algorithm and start from scratch. For NHPR-C, however, we improve each algorithm over an extended period and use only the final version of the algorithm. In this case, linkage is based primarily on identifying the same individuals through similar birthdates among two or three members of the same family across sources.

### 3.3 LINKAGE RATIO AND QUALITY OF LINKS

The most compelling evidence of correct linkage is that the resulting set of connections forms a plausible life course (see, for example, [Mayer, 2015](#)). We observe that recorded individual characteristics — such as names, birth information, partners, addresses, and occupations — often vary across sources. In principle, church records and censuses are complete, meaning that major life events — such as birth and confirmation with parental information, marriage (often including the father's name of the bride and groom), death, and census registration — should be recorded. This completeness facilitates more accurate linkage.

The current linkage rate is sufficiently high that further linking now often contributes to reducing error rates. We aim to link all individuals to their person record in all censuses. If a person P is not linked to a record in census C, the high linkage rate makes it more likely that the record "P in C" has been erroneously linked to another individual. Thus, increasing the linkage ratio from its current level often involves removing errors. Two years ago, the record "P in C" would most likely not have been linked at all.

The register changes continuously, increasing the number of links and improving quality. Continuous updates mean that versions of the NHPR must be defined by date, much like references to online sources. A 25-page report containing detailed linking statistics from the NHPR is available at [histreg](#) and is updated regularly.<sup>3</sup>

Most research projects will be based on an extract from the NHPR taken at a specific point in time; in other words, each project typically uses a static version of the register. Some researchers may prefer to add links or sources that are critical to their study before downloading their extract.

Differences in data quality and source completeness also imply substantial variation in the maximum achievable linkage rate across the closed and open periods. We expect to achieve close to 100% linkage with closed data in NHPR-C, and the linkage will be stable. This is partly because high-quality sources are available from the outset — including birthdates for several family members — and partly because there will be almost no subsequent activity aimed at improving it. Only to get access to the data to improve the linking will require time, resources and quite much planning.

By contrast, linkage in NHPR-O will not be stable, as ongoing crowdsourcing and algorithms applied by numerous researchers will continually improve linkage. Linkage rates will be considerably lower, as NHPR-O contains more than 70 million person records in sources, many with insufficient information to identify an individual.

---

3 <https://home.nr.no/~holden/HBR-statistikk.pdf>

## 4 THE OPEN REGISTER (NHPR-O)

The NHPR-O, available on the website histreg documents the deceased population using publicly accessible sources. Each individual has a profile page listing all sources in which they appear, with direct access to the corresponding entries in the Digital Archive, ensuring close integration between the register and the transcribed source material (see Figure 2). Profiles also display family relationships and biographical information, and users may add notes or explanations regarding specific linkages. There are 6.4 million such pages with at least two source entries and more than 70 million such pages with only one source.

Genealogical research is the most common activity on histreg, although many users extend their work to linking entire municipalities. Among the roughly 700 historical municipalities, more than 200 show higher linkage rates than the rest due to manual contributions — particularly in smaller municipalities. Since more than 90% of links are created by algorithms at the national level, there are no municipalities with notably low linkage rates.

Figure 2 The upper left part of the person page of Fridtjof Nansen

**Fridtjof Nansen**

Permanent ID: [pd00000033587524](#)

Mann  
Født: 10.01.1861, Aker, Akershus  
Død: 07.05.1930, Bærum

Søk etter lignende personer

**Foreldre og søsken**

Far: [Baldur Fridtjof Nansen](#), 1817 - 1885  
Mor: [Adeleide Johanne Thekla Isidora Nansen, f. Wedel Jarlsberg](#), 1821 - 1877  
Søsken: [Hans M. Nansen](#), 1854 -  
Søsken: [Mina Nansen](#), 1860 - 1860  
Denne pers.: [Fridtjof Nansen](#), 1861 - 1930  
Søsken: [Alexander Christinius Christinius Nansen](#), 1862 - 1945  
Mor: [Adeleide Johanne Thekla Isidora Nansen, f. Wedel Jarlsberg](#), 1821 - 1877  
Søsken: [Jacob Christian Kjalmar Bølling](#), 1844 -  
Søsken: [Einar Leopold Nikolai Bølling](#), 1846 -  
Søsken: [Hedevig Karen Ida Hultfeldt, f. Bølling](#), 1849 -  
Søsken: [Emil Theodor Harald Bølling](#), 1851 -  
Søsken: [Sigrid Louise Bølling](#), 1852 -  
Far: [Baldur Fridtjof Nansen](#), 1817 - 1885  
Søsken: [Cathinca Wilhelmine](#), 1854 -

**Partner og barn**

Partner: [Eva Helene Nansen, f. Sars](#), 1858 - 1907  
Barn: [Liv Høyer, f. Nansen](#), 1893 - 1958  
Barn: [Kåre Nansen](#), 1897 -  
Barn: [Aud Irmelin Nansen Revold, f. Nansen](#), 1900 - 1977  
Barn: [Odd Nansen](#), 1901 - 1973  
Barn: [Aasmund Nansen](#), 1903 -  
Partner: [Eva Helle Sars Nansen](#), 1858 -  
Barn: [Liv Høyer, f. Nansen](#), 1893 - 1958  
Barn: [Kåre Nansen](#), 1897 -  
Barn: [Aud Irmelin Nansen Revold, f. Nansen](#), 1900 - 1977  
Partner: [Sigrund Nansen Munthe, f. Sandberg](#), 1869 - 1957

### Livsløpstabell - Oversikt over kilder med personen

Nr	Dato	Kilde	PFID	Rolle	Navn	Fødselsdato	Fødested	Bosted	Fam. stil.	Sivst.	Yrke	Far/brudgom	Mor/brud
1	17.06.1859	KLOK, 0218 0301, 1852-1859	pd00000033587524	barn	Frihthjof Nansen	29.01.1859						Baldur Frillyof Nansen	Adolaiide Wedel
2	31.12.1865	FOLK, Aker, 1865-1865	pf01038011000809		Fridtjof Nansen	1861	Aker	Vestre Aker: Store-Frøen	Søn	ug			
3	31.12.1875	FOLK, Aker, 1875-1875	pf01052039009712		Frihthjof Nansen	1861	Vestre Aker Hoveds	Aker: Frøen Store	Søn	ug	Af Forældrene		

Notes: The page shows parents, siblings, partners, children and the source entries of baptism and two censuses. The table has one line per source entry, and the columns are the most important covariates from the sources. Profiling pages at histreg are only in Norwegian.

## 4.1 AUTHORITATIVE REGISTER AND IDs

As an authoritative register, NHPR-O aims to identify individuals across as many sources as possible. Such comprehensive linkage enhances the usefulness of each source and simplifies the process of gathering additional information about each individual. When the ID in NHPR-O is cited in texts, it documents the individual mentioned and enables readers to access further information about that person.

NHPR-O primarily consists of lists of PSIDs corresponding to the same individual and lists of family relationships between individuals mentioned in different sources. One PSID is designated as each person's main identifier, denoted PID, and serves as the person's ID in NHPR-O. NHPR-O employs a star-linkage approach, in which each PSID is linked to the PID for that individual. When the linkage of an individual is uncertain, users are advised to cite the PSID they are most confident corresponds to the correct individual. A PSID points uniquely to a person. But if there is uncertainty in the linking, it may be best to use several different PSIDs as reference to what most likely is the same person. This system remains robust even when links between source entries are added or removed, as all PSIDs remain well defined over time. Each PSID always refers to a specific record, regardless of subsequent merges or splits of linked records.

In some cases, it may be difficult to determine with certainty whether PSIDs referring to the same person at different stages of life — for example, during childhood with parents and siblings and later as a married adult with children — should be linked. Figure 3 illustrates the use of IDs when linkages change. Person A has the orange circle as the PID at time  $T_1$ , but all four linked PSIDs identify the same individual. In general, person A should be represented by the PID. However, if the linkage is uncertain, it is advisable to cite the PSID that most likely corresponds to the correct individual. If the linking is later modified, this ensures that the correct person can still be identified.

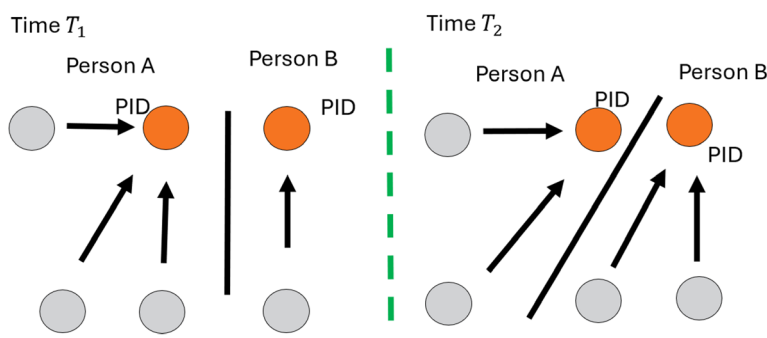
If all records referring to the same individual have not been linked, the individual will appear on multiple separate profile pages, each containing distinct groups of records or single records.

NHPR-O thus provides researchers and archive owners with both a computational system and an organizational framework that simplify the retrieval of information about each individual and provide a method for documenting their work — potentially contributing to ongoing linkage efforts within NHPR-O.

## 4.2 LINKAGE STATUS

The website histreg has been publicly available since 2016, yet almost half of the links have been created since April 2024. The underlying data system has been developed at roughly 50% of a full-time position since 2015 and includes a wide range of functionalities. In addition, substantial time has been devoted to developing algorithms for linking person records and for quality checking the links.

Figure 3 Use of IDs when linkages change



Notes: The linking is changed from time  $T_1$  to time  $T_2$ . Each circle represents a PSID and the orange circles denote the PID. The linking of the middle PSID in the lower line is shifted from person A to person B. If we want to identify this person and is not confident in the linking, then this PSID should be used in references.

Table 2 *Linking status in the open part of the NHPR in March 2026*

Census	1801	1865	1875	1891	1900	1910	1920
1801	-	18%	15%	17%	0%	0%	0%
1865	19%	-	71%	59%	54%	57%	55%
1875	20%	70%	-	54%	48%	48%	48%
1891	8%	59%	56%	-	63%	61%	59%
1900	0%	54%	49%	63%	-	76%	72%
1910	0%	57%	49%	62%	76%	-	85%
1920	0%	56%	49%	60%	72%	86%	-

*Note: The numbers show the linkage rate between the censuses where data in each row is based on the census named in the first row.*

Table 2 presents the linking status between all the open national censuses as of March 2026. Values below the diagonal represent linkage rates to earlier censuses, while values above the diagonal indicate linkage rates to later censuses. The matrix would be symmetric were it not for cases in which individuals are registered twice in a census or are absent because they are abroad. Of the 13.4 million person records in the nationwide censuses, 10.1 million (75%) are linked to other source entries. The linkage rates demonstrate broad coverage across the population, regardless of region, birth cohort, sex, or occupation.<sup>4</sup>

The NHPR-O is widely accessible on the website histreg, but this availability also presents some challenges. In October 2025, histreg received more than 200,000 visits per hour, challenging the computer system at the National Archives. Most of this traffic is not generated by manual users. Rather, it is dominated by search engine indexing and the use of the site as training data for artificial intelligence. It is therefore difficult to quantify visits by users interested in genealogy or research. We also observe that downloads of the paper by Thorvaldsen and Holden (2023), which is used as documentation for histreg on the website, increased after search engines began indexing the site. It has become necessary to impose restrictions on visits and to encourage dialogue with users who have special requests. Also, after introducing technology verifying a human user the number of visits still challenge the computer system of the National Archive. Unfortunately, this also implies that profiling pages at present are not indexed by search engines. This illustrates the importance and complexity of availability at internet.

## 5 THE CLOSED REGISTER (NHPR-C)

The NHPR-C uses the same data structure as the NPR. This alignment is necessary to facilitate integration with modern register data. Privacy requirements, varying source types, and the processes of transcription and linkage necessitate a distributed data structure that differs from that of NHPR-O. Data from NHPR-O are copied to NHPR-C at regular intervals so that NHPR-C covers the total population across the entire period as described in Table 1 and the notes to Figure 1. When new sources become open, these records are copied from NHPR-C into NHPR-O. The Danish Multigeneration Register (Due et al., 2024) identifies the family relations from 1920 to the present which is a central part of building NHPR-C. The Danish register is released in May 2026.

The integrated NPR and NHPR-C include individuals born after 1964 who are fully described in the NPR; individuals, both living and deceased, who are registered in both open and closed sources; and deceased individuals from the 19th and early 20th centuries who are registered only in open sources.

The data structure contains a name table with all permanent information for each individual — such as IDs, birth and death details, and parents — as well as tables for time-varying data, including name changes, household membership, partners, adoption, and similar events.

<sup>4</sup> <https://home.nr.no/~holden/HBR-statistikk.pdf>

NHPR-C, including the NPR, extends the microdata infrastructure back to 1801, increasing the genealogical depth from two to seven generations. This requires the use of privacy-friendly distributed linkage and data minimisation when data are exported to research projects, like the methods routinely applied to modern classified data. All data, including 19th-century records, must also be pseudonymised and minimised when used together with modern data. This prevents living individuals from being identified through their ancestors.

## 5.1 IDs IN NHPR-C

Each individual in NHPR-C may have up to three IDs: the birthdate-based ID used in the current NPR and used daily by all Norwegians in contexts such as health, finance, and employment; an NHPR-C-C ID used in closed sources; and an NHPR-C-O ID used in open sources. Thus, NHPR-C introduces two new IDs with the extensions "-C" and "-O", used in closed and open sources respectively.

Each source has an ID table connecting the PSID to either the current birthdate ID, the closed NHPR-C-C ID, or the open NHPR-C-O ID, depending on whether the source already uses the birthdate ID and whether the source is open or closed.

The NHPR-C-C ID used in closed sources is non-informative, making it impossible to obtain additional information about an individual by accessing a single restricted source. Misuse would require access to at least two closed sources. The NHPR-C-O ID corresponds to one of the linked PSIDs and follows the same principles as IDs in NHPR-O. These IDs are flexible and remain well defined regardless of later merges or splits of linked individuals.

## 5.2 A TYPICAL DATA-MINIMIZED DATA SET

NHPR-C does not record whether an individual appears in a specific source, such as the 1950 census. An individual living in Norway in 1950 is very likely registered in that census, but this information is stored only within the census itself, not in NHPR-C. The same principle applies to all sources.

While it is not sensitive to know that an individual appears in a census, it is highly sensitive to know that the same individual is registered in, for example, the Cancer Registry. From both the 1950 census and the Cancer Registry, an individual can be identified in NHPR-C using the NHPR-C-C ID.

For example, Statistics Norway may release a dataset containing only the following variables: year of birth and death, cause of death when available, county of residence at birth, occupational code, and project IDs for the individual, father, and mother for a large part of the Norwegian population over seven generations. Such a dataset would be a rectangular table with eight columns (covariates) and about nine million rows (individuals). This type of dataset is valuable for research on life expectancy. Access is granted only to approved research projects, with the requirement that individuals must not be identified.

This data structure allows certain datasets to be created using distributed linking, in which data from different sources are combined only after approval has been granted for each source and after project-specific IDs have been generated. As a result, a complete dataset containing all identifiable information will never exist; only pseudonymised versions will be available.

## 6 THE NHPR AS A RESEARCH INFRASTRUCTURE

The NHPR has numerous application areas. This is well documented in several papers in this journal, such as Vikstrøm et al. (2023). The main reason for funding the NHPR is academic research. However, we anticipate that its primary use will be among the public for genealogical and local history research. This activity will improve linkage ratios and data quality in the NHPR and add new sources in the coming decades. In addition, the NHPR will be a resource for search engines and AI training, which will continuously challenge the computer system and its accessibility.

Different researchers and scientific disciplines focus on different aspects of population registers. Several steps are necessary to realise the full potential of such registers: transcribing sources; linking records in the primary sources for as large a part of the population as possible while ensuring that minority populations are included; expanding the data with information from additional registers; establishing a tradition of using an authoritative register; and finally utilising the linked data and all sources across disciplines, possibly in combination with modern register data.

It is essential to build an infrastructure in which researchers from different fields both contribute to and benefit from the results, and in which each researcher and contributor is recognised for their contributions. This is most relevant for open data but may also apply to closed data. Manual improvement of the open register of public servants (1867–1960) is one example involving open sources. Two examples involving closed data are military health protocols and school protocols from boarding schools for minorities after the Second World War. To make transcription and linkage of these sources available for others serve as a documentation of own research and makes new research much easier, both for similar research questions as the original research, but also in completely different directions.

## 6.1 OPEN DATA

Open population data are very valuable for research, as demonstrated in several studies on population databases (Ruggles et al., 2025). When research is limited to deceased individuals, anonymisation is rarely required. As a result, it is easier to gain access to a wide range of sources and to verify and analyse the population, society, and each source in far greater detail.

On each individual profile in NHPR-O, variations in recorded personal data — such as name and birthdate — can easily be observed across sources. An individual may be mentioned in anywhere from 5 to 50 different sources. Holden et al. (2025) estimated that 14% of birthdates in the 1910 census and 17% of birthdates in the 1920 census were incorrect. The error rate increases with age, is higher in the northern parts of Norway, and varies by birthplace, household role, and occupation.

It is understandable that historical research to date has focused on the largest and most systematic migrations with well-documented causes. By tracking each individual in the NHPR, it is possible to analyse both systematic and more random migration between censuses and to identify the age, sex, and occupation of migrants. Holden et al. (2025) were able to quantify migration between counties and municipalities from 1910 to 1920 using NHPR data. The proportion of individuals remaining in the same county over this decade varies between 72% and 97%. Migration rates are higher in the coastal areas of each county.

By incorporating as many different sources as possible, we can construct a more detailed description of each individual. This also increases the accessibility and value of each source. For example, several historical research projects have examined the treason archive from the Second World War, which contains registrations for around 90,000 individuals and companies (e.g. Seemann, 2019). Names are open, while descriptions of each case are restricted. When these individuals and company owners can be identified in additional sources, new opportunities arise for both qualitative and quantitative research on the individuals and families involved.

Because this increases the value of each source, archive-owning organisations and researchers who analyse individual sources are motivated to link the person records in their sources to the NHPR. This, in turn, enhances the value of the source itself, the research based on it, the entire NHPR, and all sources connected to it.

## 6.2 USE OF THE AUTHORITATIVE REGISTER

Many texts, including scientific articles, refer to individuals who are difficult to identify without specialised knowledge. From a scientific perspective, this is problematic, as scholarly work requires precise and verifiable documentation. The NHPR provides a platform for identifying each individual and documenting references in the text. It also provides a platform for locating additional information, which strengthens the NHPR for everyone.

The website histreg also makes it possible to add references for each individual to sources not included in NHPR, such as lexica, local history books, or publicly available censuses from other countries. As of March 2026, more than 20,000 manually registered references are included in the NHPR. All these references are included in a reference list, making it easy to find individuals mentioned in the same source and to check the quality of references through random checks.

Its role as an authoritative register is therefore important both within and beyond scientific research. We hope it will be used more extensively in a wide range of contexts, including academic writing, museums, exhibitions, presentations, and other settings involving historical individuals.

### 6.3 CLOSED DATA

As in other Nordic countries, Norway is known for the high quality of its register data, owing to an extensive public welfare system, a relatively homogeneous population, and the NPR with its unique birthdate-based identifier. As a result, Norway has comprehensive registers covering health, education, occupation, taxation, family relations, household composition, addresses, and more for the entire population.

The last census was conducted in 2001, as the relevant information is continuously updated in public registers, making new censuses unnecessary. Between 1911 and 2007, each taxpayer submitted a form with information on income and property to the Tax Administration. Since 2008, the Tax Administration has instead sent this information to taxpayers for confirmation. This illustrates the extent of register-based data built on the NPR identification system.

The Death Register, including information on cause of death, has been complete since 1928, and the Cancer Register since the mid-1950s. Extending the NPR from two to seven generations significantly increases the value of the register data. These data are invaluable for research across academic disciplines, including the social sciences, medicine, economics, and the humanities.

### 6.4 REPRESENTATIVITY

Representativity is generally a minor concern in a register that aims to cover the entire population, as shown in Holden et al. (2025). This can be illustrated with two simple numerical examples.

When 80% or more of the population is linked, overrepresentation of some groups becomes less problematic. For instance, if 82% of men and 78% of women are linked, men travel 50% more than women, and the estimate is not adjusted for gender, the estimated number of trips will be only 0.6% too high — well within the uncertainty of such estimates.

Underrepresentation of small minorities — such as foreigners, construction workers, or itinerant groups — poses a problem only if the phenomenon being studied is very atypical within the minority. A small group will have little effect on average life expectancy but may have a large impact on indicators such as mobility between counties, child mortality, and other low-frequency events. If only 5% of a group representing 5% of the population is linked, and this group travels five times more than the rest of the population, while 80% of the majority population is linked, overall mobility will be underestimated by roughly 15%. The same reasoning applies if "mobility" is replaced with "child mortality." When such groups can be identified in the censuses, it is possible to compensate for the lower linkage rate based on the rate observed in the census and thereby avoid underestimation.

This argument also implies that all available data can be used in the linkage process, even if this results in higher linkage rates for individuals with rare occupations or names. When relying on crowdsourcing, it is necessary to allow contributors to use all available information.

Open research questions include the completeness of the sources and the maximum achievable linkage rate. Holden et al. (2025) estimate that it is possible to link 94% of individuals between the 1910 and 1920 censuses. This estimate is based on 200 individuals randomly selected from the 1920 census among the 20% of the population that had not already been linked between the censuses when the analysis was performed. One third of the remaining 6% were most likely outside Norway in 1910. Higher linkage rates are required to draw robust conclusions.

Related research questions concern the extent to which individuals are registered twice in censuses without being marked as non-permanent residents, and how many individuals are entirely missing from the records. For example, 3.4% of the population in the 1900 census is registered as not present or as non-permanent residents. These individuals should therefore be registered twice in the census. Current linkage results indicate that a similar number of individuals appear to be registered twice without such notation. This suggests that the population may be about 3.4% smaller than expected.

There are five main groups with double registration in the census:

1. Individuals, mainly male, working or studying far from their home or family
2. Young individuals, more often female, in the process of leaving their parents' home
3. Individuals staying in hospitals
4. Children partly cared for by people other than their parents
5. A limited number of wealthy individuals and politicians with several addresses

The first group is the largest.

The analysis above indicates that 4% of the population is difficult to find in the 1910 census. It is too early to quantify how much this number may be reduced. But it is reasonable to believe that at least 1% of the individuals are missing from sources in which they should have been recorded. Missing registrations imply that the census underestimates the population. The reliability of this estimate depends critically on the quality of the transcribed censuses. Birth year is missing for 0.03%–1.3% of individuals in the transcribed national censuses. The censuses also show that about 20% of individuals do not live with their families. These individuals are more difficult to identify because we have data for only one individual and the information is generally poorer. The proportion with missing birth year is up to four times higher in this group.

## 7 INTERNATIONAL COOPERATION

### 7.1 MIGRATION DATA

Emigration lists are included in the NHPR and cover the extensive Norwegian emigration to America. These lists are believed to be almost complete, but they contain so little information about each individual that identification is often difficult. For this reason, it would be valuable to include individuals born in Norway who appear in US censuses as an additional source in the NHPR. As of March 2026, there are more than 300 references to US-based sources, registered manually with links to sources at FamilySearch and Find a Grave.<sup>5</sup> To access the data in FamilySearch, users are required to log in. This documentation is acceptable, but preferably these references and links should be registered using a method that requires less manual work and is more transparent.

Registration of immigrants is rare. The proportion of immigrants increases from 1.2% to 2.8% between 1865 and 1920 based on birthplace information in the censuses. Individuals from Sweden accounted for 65–80% of all immigrants. In the 1920 census, the share of individuals born abroad ranged from 1% to 8% across counties.

By combining the NHPR with birthplace information in the censuses, it is possible to compile lists of individuals who immigrated to Norway, as well as those who emigrated from Norway. The NHPR project is interested in exchanging lists of immigrants and emigrants with the owners of similar databases in other countries to track individuals and their families as accurately as possible — ideally in ways that allow the use of crowdsourcing in the linkage process.

The exchange of migration data makes it possible to gradually extend the register, moving towards an international population register. This will open new avenues for research, such as comparing populations across countries or comparing immigrants with siblings who remain in Norway. Simply increasing the size of the population under study is invaluable for many types of research, particularly

5 <https://www.familysearch.org/en/norge/>; <https://www.findagrave.com/>

in medicine. Most rare diseases are genetic, and there are documented cases in which genetic diseases have been traced across generations, including in families that have migrated between countries. See Sokka (2004) for examples of tracking genetic diseases.

## 7.2 INTERNATIONAL POPULATION REGISTER

There are several examples of international cooperation on population registers. IPUMS International, for instance, distributes census data from 104 countries, with access restricted to research purposes (Ruggles et al., 2025). The use of this and similar databases in numerous scientific articles demonstrates the value of international population registers. If such databases were extended to the present and made available alongside modern register data, their value would increase manyfold. That said, the privacy and access regulations governing the different databases also pose significant challenges.

Privacy and legal restrictions are more significant and challenging for individuals who are still alive. In Europe, a common legal framework applies through the GDPR, but national interpretations differ in the details of its implementation (Molnár-Gábor et al., 2022). In general, the handling of information about living individuals is strictly regulated. Information about deceased individuals is typically not considered personal data; however, such data may still be subject to legal restrictions depending on how it was collected and on the ownership of the source and its transcription. Medical data concerning deceased individuals, including causes of death, may also be considered personal data with respect to their living relatives. It is common practice in European countries to conduct register-based research on living individuals using pseudonymisation and data minimisation.

We anticipate that countries will not permit the export of personal data to other countries. An international population register that includes the present population must therefore be distributed, with separate databases maintained within each country. This makes it necessary to develop international rules for person identifiers that cover the entire period from the beginning of each population register to the present. Nevertheless, each country may continue to use its own national person-ID system.

A separate register is also needed for individuals who migrate between countries, including the person IDs assigned to them in both the emigration and the immigration country. This register may initially be established bilaterally between emigrating and immigrating countries, but it should share the same core properties. It is also necessary to handle individuals who continue to migrate to additional countries. This task is easier for countries with an active national population register, as in the Nordic countries. Countries without a national population register, such as Germany, would first need cooperation between different regions and harmonisation of regional systems before participating in an international exchange of data.

Slightly fewer than one million of the 5.5 million people living in Norway today were born abroad. Many other countries have similar or even higher numbers of immigrants. About 60% of these immigrants come from countries outside Western Europe and North America, where high-quality population registers or data exchange with Norway cannot always be expected. As a result, a large minority of the population will not be included in a multigenerational population register.

Building an NHPR in each country and establishing international cooperation between NHPRs are major undertakings that may take decades to complete. It is important to design the register so that each new step in its development provides additional possibilities and a return on investment. The register can be constructed in such a way that each new source added, and every new link created, enables a larger share of the population to be followed. This, in turn, makes new types of research possible, and the overall value of the register increases with each step.

Data exchange between countries concerning recent migration involving individuals who are still alive is likely to be a more politicised issue than a purely research-related matter. However, this should not prevent progress on the initial and more feasible steps towards international cooperation between NHPRs. Substantial value can be achieved even without incorporating the most recent migrations.

## 8 CONCLUDING REMARKS

The NHPR covers the entire Norwegian population from 1801 to the present. It follows the same principles as similar historical databases but also has several additional characteristics: linkage to modern microdata, an authoritative register, the inclusion of many sources, and the use of both crowdsourcing and algorithmic linkage. The introduction of an authoritative register encourages better documentation in all texts and presentations that mention historical individuals from the broader population. The paper also discusses how a national population register may be extended into an international register.

The NHPR opens new research possibilities across many disciplines. Some of these require new methods, such as handling privacy when the register includes both present and historical populations and developing a flexible ID system that accommodates changes in the linking of records across sources. The paper also discusses fundamental research issues such as representativity and the proportion of individuals who are either registered twice or missing from primary sources.

## ACKNOWLEDGEMENT

The authors thank The National Archives of Norway, Statistics Norway, Norwegian Institute of Public Health, UiT – the Arctic University of Norway and all contributors to histreg. Further, the authors are grateful to Gunnar Thorvaldsen, Kåre Bævre and Lars Nygaard and the Research Council of Norway under contract 322231.

## REFERENCES

- Alter, G., Mandemakers, K., & Vézina, H. (2023). Introduction: Content, design and structure of major databases with historical longitudinal population data. *Historical Life Course Studies*, 13, 228–234. <https://doi.org/10.51964/hlcs15759>
- Due, J. K., Pedersen, M. G., Antonsen, S., Rommedahl, J., Agerbo, E., Mortensen, P. B., Sørensen, H. T., Lotz, J. F., Piqueras, L. C., Fierro, C., Karamolegkou, A., Igel, C., Rust, P., Søgaard, A., & Pedersen, C.B. (2024). Towards more comprehensive nationwide familial aggregation studies in Denmark: The Danish Civil Registration System versus the lite Danish Multi-Generation Register. *Scandinavian Journal of Public Health*, 52(4), 528–538. <https://doi.org/10.1177/14034948221147096>
- Holden, L., & Boudko, S. (2018). The Norwegian Historic Population Register and migration. *Journal of Migration History*, 4(2), 249–263. <https://doi.org/10.1163/23519924-00402002>
- Holden, L., Boudko, S., & Fjellvoll, B. (2025). Historisk befolkningsregister som et autoritetsregister for personer og verktøy for lokalhistorisk forskning [Norwegian Historical Population Register as an authority register for persons and a tool for local historical research]. *Heimen*, 62(4), 311–328. <https://doi.org/10.18261/heimen.62.4.4>
- Holden, L., Boudko, S., & Thorvaldsen, G. (2020). Lenking og kobling i Historisk befolkningsregister [Record linkage and family-relations in the Norwegian Historical Population Register]. *Heimen*, 57(3), 216–229. <https://doi.org/10.18261/issn1894-3195-2020-03-04>
- Mandemakers, K. (2025). Overview and comparison of 84 databases with historical population longitudinal micro data. *Historical Life Course Studies*, 15, 281–321. <https://doi.org/10.52024/hlcs21660>
- Mandemakers, K., & Dillon, L. (2004). Best practices with large databases on historical populations. *Historical Methods*, 37(1), 34–38. <https://doi.org/10.3200/HMTS.37.1.34-38>
- Mayer, K. U. (2015). An observatory for life courses: Populations, countries, institutions, and history. *Research in Human Development*, 12(3–4), 196–201. <https://doi.org/10.1080/15427609.201.1068051>

- Molnár-Gábor, F., Sellner, J., Pagil, S., Slokenberga, S., Tzortzatou-Nanopoulou, O., & Nyström, K. (2022). Harmonization after the GDPR? Divergences in the rules for genetic and health data sharing in four member states and ways to overcome them by EU measures: Insights from Germany, Greece, Latvia and Sweden. *Seminars in Cancer Biology*, 84, 271–283. <https://doi.org/10.1016/j.semcancer.2021.12.001>
- Ruggles, S., Cleveland, L., Lovatón Dávila, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Heimann, Q., Lee, J., & Merrill, N. (2025). *Integrated Public Use Microdata Series, International: Version 7.6* [Dataset]. IPUMS. <https://doi.org/10.18128/D020.V7.6>
- Seemann, A. (2019). Citizen outcasts: The penalty of 'loss of civil rights' during the Norwegian treason trials, 1945–1953. *Scandinavian Journal of History*, 45(3), 360–383. <https://doi.org/10.1080/03468755.2019.1620327>
- Sokka, T. (2004). National databases and rheumatology research I: Longitudinal databases in Scandinavia. *Rheumatic Disease Clinics of North America*, 30(4), 851–867. <https://doi.org/10.1016/j.rdc.2004.07.002>
- Thorvaldsen, G., & Holden, L. (2023). The development of microhistorical databases in Norway. A historiography. *Historical Life Course Studies*, 13, 127–147. <https://doi.org/10.51964/hlcs14315>
- Vikström, P., Larsson, M., Engberg, E., & Edvinsson, S. (2023). The Demographic Database — History of technical and methodological achievements. *Historical Life Course Studies*, 13, 89–102. <https://doi.org/10.51964/hlcs12163>