

Norwegian Historical Population Register, 1801-present

Lars Holden,¹ Svetlana Boudko,¹ Bjørn Fjellvoll,¹

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.

1. Introduction

The Norwegian Historical Population Register (NHPR) covers the entire Norwegian population between 1801 to present. It consists of a public register of deceased individuals documented in open sources, available at histreg.no, and a closed register that extends the National Person Register (NPR) and its identification system used daily by all Norwegians. As a result, the NHPR may be combined with modern microdata. The main purpose of the NHPR is to serve as a research infrastructure across disciplines, although it is also widely used in genealogical work.

Alter et al. (2023) and Mandemakers (2025) offer overviews of different longitudinal population databases, and the NHPR follows many of the same principles as similar

¹ Norsk Regnesentral, <https://nr.no/>

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 immigrations to Norway are typically registered only when individuals appear in church records or censuses. This is previously studied in Holden and Boudko (2018).

The NHPR serves as an authoritative register of deceased persons, assigning a unique identifier (ID) to each individual. This ID is intended to be used 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 become 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 person, thereby increasing research value and enhancing the utility of all linked sources.

Most primary sources have been transcribed. The open sources are available in the Digital Archive (Digitalarkivet.no), which is managed by the National Archives of Norway. The Digital Archive assigns each individual named in a source a unique and permanent identifier, denoted PSID. The NHPR's identification system is based on these PSIDs via an API. Census data are subject to a 100-year privacy restriction, while church records have a 60-year limit.

During the initial phase of the NHPR-project, substantial effort was devoted to supporting the transcription of primary sources. Some of the sources, including the closed 1930 and 1950 censuses, are being transcribed using machine learning. Transcription quality improves during linking with repeated information, such as birthdates within families.

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. 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).

2. Linking Methods

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

Large volumes of the links are straightforward, for example the majority of 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 also many links that only can only be established manually.

Manual judgement is necessary when linking records with sparse information or unusual variations. A limited ratio of the links is 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 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, but where it is unclear whether a particular record in the emigration list is the correct individual.

Several different algorithms are used in finding links between person records in the sources. Most of the algorithms try to link each person in a source sequentially from the first person record in the source according to certain criteria. The algorithms identify persons and families with similar characteristics (names, birth date/year, municipality etc.) in different sources, systematically between or within specific sources or in an open search in all sources in the Digital Archive. The algorithms also consider the present linking of the person. The algorithms for quality check systematically review all profile pages and identify large deviation in characteristics, several registrations in the same source, (census, birth or death) or unlikely life courses. The algorithms for quality check automatically remove erroneous links or flagging linked persons for manual inspection. This creates a structured and dynamic interaction between automated processes and manual contributors.

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 neglecting family relationships from church books since in some cases stepparents may be registered as parents.

More than 260 individuals have contributed to the crowdsourcing by registering links. This implies that the database is continuously growing and improving. These contributors are highly experienced and possess extensive knowledge of genealogy and the relevant sources. All links are registered with contributor and timestamp. The timestamps show that the activity spreads across nearly all hours of the year. This is like putting together a puzzle on the internet. It's easy to get addicted.

Histreg 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 representativity and data quality while increasing contributors' efficiency. It enables the contributors to focus on person records that need manual inspection in municipalities and time periods where the contributor has particular interest or expertise.

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 almost never removed, this requires stricter rules for each new link, making it more challenging to obtain high linkage ratios. While no formal statistical evaluation has been conducted, our experience indicates that manual contributors generally achieve lower error rates than algorithms. This knowledge is based on a test period where we evaluated the contributor/algorithm that have made links that are flagged as erroneous.

The most compelling evidence of correct linkage is that the resulting set of connections forms a plausible life course. We observe that recorded personal characteristics—such as names, birth information, partners, addresses, and occupations—often vary across sources. In principle, the church records and censuses are complete, meaning that all major life events—such as birth, confirmation, marriage, parental information, often the father's name at marriage, death, and census registration—should be documented. This completeness facilitates more accurate linkage. The current linkage rate is sufficiently high that further linking now contributes to reducing error rates. A missing registration in a primary source may be incorrectly linked to another person. Two years ago, a missing registration would most likely not have been linked at all.

The register changes continuously, increasing number of links and improves the quality. Continuous updates mean that versions of the NHPR must be defined by date, much

like references to online sources. A 20-page report containing detailed linking statistics from the NHPR is available at histreg.no and is updated regularly.²

Most research projects will be based on an extract from the NHPR taken at a certain point in time; in other words, each project typically use 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 for the restricted period. In the open period, however, linkage rates will be considerably lower, as it contains more than 100 million person records in sources many with insufficient information to identify a person. Linkage in the closed period will be stable. First, because we obtain high quality from the outset, given the availability of high-quality sources that include birth dates of several family members; and second, because there will be almost no subsequent activity aimed at improving it. By contrast, the open period will not be stable, as ongoing crowdsourcing and algorithms applied by numerous researchers will continually improve linkage.

3. The Open Register (NHPR-O)

The NHPR-O, available at the website histreg.no, 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 to ensure close integration between the register and the transcribed source material.³ Profiles also display family relationships and biographical information, and users may add notes or explanations regarding specific linkages.

New transcribed sources are continuously added, including thematic registers from 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. The NHPR-O also incorporates 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, and material extracted from newspapers and biographical sources. All individuals mentioned in these sources are assigned PSIDs and may be linked to the NHPR-O in the same way as an entry derived from church records or censuses.

The 1920 census and a list of deaths up to 2014 are the most recent sources included. A person may be mentioned in anywhere from five to fifty different sources. If all records referring to the same individual have not been linked, the person will appear on multiple

² <https://home.nr.no/~holden/HBR-statistikk.pdf>

³ A typical profile page: <https://histreg.no/person/pf01073681015788>

separate profile pages, each containing distinct groups or single records. The profile page makes it easy to observe how names, birth data, and other characteristics vary across sources.

The open population for 1800–1920 includes approximately 6.6 million individuals across around 70 million source entries. As of December 2025, 25.1 million persons named in sources are linked to 6.0 million persons. For 1801–1963, the population numbers 9.1 million individuals.

Histreg.no has been publicly available since 2016, yet more than one-third of all 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 array of functionalities. In addition, substantial time has been devoted to developing the algorithms used for linking person records in sources and quality checking the links.

Genealogical research is the most common activity on histreg.no, though 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 generated by algorithms at the national level, there are no municipalities with notably low linkage rates.

As an authoritative register, the 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 details about them.

The NHPR-O primarily consists of lists of PSIDs corresponding to the same individuals and lists of family relationships between persons mentioned in different sources. One PSID is designated as each person's main identifier and the person's ID in NHPR-O. The NHPR-O employs a star-linkage approach, in which each PSID is linked to the main PSID for that individual. When the linkage of a person is uncertain, users are advised to cite the PSID they are confident corresponds to the correct person. This system remains robust even when links between source entries are added or removed, as all person IDs remain well-defined over time. Each PSID always points to a specific individual, regardless of subsequent mergers or splits of records.

It should be emphasised that the NHPR-O is a register of individual records derived from archives and sources and is not intended to replace the various archives that document different aspects of Norwegian history. We aim to collaborate with the owners of the different archives by enhancing the value of their collections and directing more visitors to their websites as part of our shared effort to make the NHPR-O as comprehensive as possible.

The NHPR-O thus provides researchers and archive owner with both a computational system and an organisational framework that simplify the retrieval of information on each person and offer a method for documenting their own work—potentially contributing to ongoing linkage efforts within the NHPR-O.

Table 1 presents the linking status as of December 2025. 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 due to being abroad. Of the 13.4 million person records in the nationwide censuses, 9.7 million are linked to other source entries. The linkage rates demonstrate broad coverage across the population, regardless of region, birth cohort, sex, or occupation.

Census	1801	1865	1875	1891	1900	1910	1920
1801	--	18 %	12 %	9 %	0 %	0 %	0 %
1865	19 %	--	64 %	48 %	45 %	44 %	44 %
1875	10 %	62 %	--	48 %	43 %	41 %	41 %
1891	14 %	48 %	50 %	--	60 %	54 %	55 %
1900	0 %	45 %	44 %	61 %	--	72 %	68 %
1910	0 %	44 %	42 %	56 %	72 %	--	85 %
1920	0 %	45 %	42 %	55 %	69 %	86 %	--

Table 1. Linking status in the open part of the NHPR in December 2025. The number shows the linkage rate between the censuses where data in each row is based on the census named in the first row.

4. 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 transcription and linkage processes necessitate a data structure that differs from the NHPR-O's. Data from NHPR-O will be transferred to NHPR-C at regular intervals.

The integrated NPR and NHPR-C include persons born after 1964 that 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 nineteenth and early twentieth centuries who are only registered in open sources.

The data structure contains a name table with all permanent information for each person—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.

Each person in the NHPR-C may have up to three IDs: the birth-date-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. Each source has an ID table connecting the PSID to either the present birthdate ID, the closed NHPR-C-C ID, or the open NHPR-C-O ID, depending on whether the source already uses the birth-date 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 a person by accessing a single restricted source. 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 and splits of linked individuals.

The NHPR-C does not record whether a person appears in a specific source, such as the 1950 census. A person living in Norway in 1950 is very likely registered in the 1950 census, but this information is stored only within the census itself, not in the NHPR-C. The same principle applies to all sources. While it is not sensitive to know that a person 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, a person can be located in the NHPR-C by using the NHPR-C-C ID.

There are more primary sources in the closed period. A paper-based population register documenting all changes within a family was introduced in the capital, Oslo, in 1906, later extended to other large cities, and implemented nationwide between 1946 and 1995. For the years in which it overlaps with the NPR, this register contains more detailed information on addresses and parents. There is also a paper-based registration of births, marriages and deaths for the period 1908–1960. Both sources contain birth dates for several family members, which are invaluable for linkage.

The integrated NHPR-C and NPR extends the microdata infrastructure 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 is exported to research projects, like the methods routinely applied to modern, classified data. All data, including the 19th-century records, must also be pseudonymised and minimised when used together with modern data. This prevents living individuals being identified through their ancestors.

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 family relationships between individuals for large parts of the Norwegian population over seven generations. Such a dataset could be used for research on life expectancy. Access is granted only to approved research projects, with the requirement that the individuals are not to be identified.

This data structure allows for the creation of certain datasets 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.

The first version of the NHPR-C will be completed at the end of 2025 and will be gradually improved throughout 2026.

5. 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 for academic purposes. However, we anticipate its primary use will be among the public for genealogical and local history research, or as a resource for search engines and AI training. This use will improve the linking ratios and quality in NHPR and add new sources in the coming decades.

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 birth-date-based identifier. As a result, Norway has comprehensive registers covering health, education, occupation, taxation, family, 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 to 2007, each taxpayer submitted a form with information on income and properties to the Tax Administration. Since 2008, the Tax Administration has instead sent this information to taxpayers for confirmation. This shows the extent of register-based data built on the NPR's identification system.

The Death Register, including information regarding 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, but not limited to the social sciences, medicine, economics, and the humanities.

Even without the modern register data, the NHPR is a valuable resource 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 becomes possible to use a wide range of sources and to verify and analyse the population, society, and each source in far greater detail. Access to data is also considerably easier and less bureaucratic.

On each individual profile in the NHPR-O, variations in recorded personal data—such as name and birthdate—can easily be observed across 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, 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 the 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 Second World War, which contains registrations for around 90,000 individuals and companies (e.g. Seemann, 2019). When these individuals and company owners can be identified in additional sources, new opportunities arise for both qualitative and quantitative research on the persons 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.

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

Histreg.no 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 December 2025, more than 18,000 manually registered references are included in the NHPR. 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 persons.

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 the records in the primary sources for as large part of the population as possible, ensure that each minority population is included; expanding the

data with information from additional registers; establishing a tradition for the use of an authoritative register; and, finally, utilising the linked data and all the 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 recognized for their contributions.

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 of number of trips will only be 0.6 % too high—well within the uncertainty of such estimates.

Underrepresentation of small minorities—such as foreigners, construction workers, or itinerant groups—only poses a problem 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 like mobility between counties, child mortality, and the occurrence of 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 are 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 and thereby avoid underestimation. This argument also implies that all available data can be used in the linkage process, even though this may result in higher linkage rates for individuals with rare occupations or names. When relying on crowdsourcing, it is necessary to allow contributors to use all types of available information.

Open research questions include the completeness of the sources and the maximum achievable linkage rate. Higher linkage rates are required to draw robust conclusions. Statistical analyses and the highest linkage rates observed in ten different municipalities suggest that it is possible to achieve a linkage rate of approximately 95% for person records in censuses in municipalities with limited migration and outside the largest cities between the 1910 and 1920 census, Holden et. al. (2025). Related research questions include 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. As an example, 3.4% of the population in the 1900 census is registered as not present or non-permanent residents. These persons should be registered twice in the census. The linking so far indicates that there is similar number that are registered twice, without such registration. This indicates that

the population may be 3,4 % smaller than expected. There are mainly three groups with double registration in the census, persons, mainly male, working far from home/family, young persons, more often female, in the process of living parents' home and children that are (partly) taken care of by other persons than the parents. The first group is the larger of the three groups.

More than 1% of individuals cannot be identified in censuses where they should be registered. But it is still too early to quantify how many individuals are missing from sources where they should have been registered. Missing registrations imply that the census underestimates the population. The quality of this estimate depends critically on the quality of the censuses. Birthyear is missing for 0.03%-1,3% of the persons in the transcribed national censuses. About 20% of the persons don't live with their family. These persons are more difficult to identify, since we have data for only one person and, in general, the data is poorer. The ratio for missing birthyear is up to 4 times higher for this group.

The NHPR-O is widely accessible on the website histreg.no, but this availability also presents some challenges. In October 2025, histreg.no had more than 200.000 visits per hour, challenging the computer system in the National Archive. Most of this traffic isn't manual. Rather, it is dominated by search engine indexing and as a training data for artificial intelligence. It is difficult to quantify visits by those who interested in genealogy or who are using it for research. It is necessary to impose restrictions on visits and encourage dialogue with users who have special requests. We also observe that downloads of the paper by Thorvaldsen and Holden (2023), which is used as documentation of histreg on the website, increased after search engines began indexing histreg.⁴ This illustrates the importance of availability and highlights the difficulty of identifying and predicting which actions are most critical for visibility.

6. International cooperation

Emigration lists are included in the NHPR and cover the extensive Norwegian emigration to America. These lists are 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. There are more than 200 references, by December 2025, to sources in US based on manual registration with links to sources at FamilySearch.⁵ In order to read the data in the source, it is required to log in at FamilySearch. This documentation is fine. But, preferably, these references/links should be registered with a method that requires less manual work and is more transparent.

⁴ <https://hlcs.nl/article/view/14315>

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

Registration of immigrants is nearly non-existent. Individuals from Sweden accounted for 65–80 % of all immigrants, based on birthplace information in the censuses between 1865 and 1920. In the 1920 census, the share of individuals born abroad ranged from 1 % to 8 % across counties.⁶ 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 migrants and emigrants with the owners of similar databases in other countries to track each individual and their family as accurately as possible—ideally in ways that allow for 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 of research, such as comparing populations across countries or those who migrate with siblings who remain in Norway. Simply increasing the size of the population under study is invaluable for many types of research, particularly 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 Solla (2004) for tracking of genetic diseases.

There are several examples of international cooperation on population registers. The 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 demonstrate 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.

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:

- OPEN: fully open and eligible for publication on the internet
- RESTRICTED: researchers may obtain full access to the sources
- CLOSED: researchers may access pseudonymised data, subject to data minimisation requirements

Privacy and legal restrictions are more significant and challenging for individuals who are still alive. In Europe, a common legal framework applies through GDPR, but national interpretations differ in the details of its implementation (Molnár-Gábor et al., 2022). In general, the handing of information about living persons is strictly regulated.

⁶ Based on birthplace in censuses

Information about deceased persons 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 cause of death, may be considered personal data for their living relatives. It is common practice in European countries to conduct register-based research on living persons 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 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 both in the emigration and the immigration country. This register may be established bilaterally between emigrating and immigrating countries, but it should share the same core properties. We also need to handle persons that continue to migrate to new countries. This is easier for countries with an active national population register, as in the Nordic countries. Countries that do not have an NPR, as in Germany, would first need cooperation between different regions and harmonisation between 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 we cannot expect high-quality population registers or data exchange with Norway. 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 still alive is likely to be a more politicised than purely the topic of a research infrastructure. 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.

7. Concluding remarks

The NHPR covers the entire Norwegian population between 1801 to present. It follows the same principles as similar historical databases but also has other characteristics as: link to modern microdata, an authoritative register, includes many sources and both crowdsourcing and algorithmic linkage. The introduction of an authoritative register encourages better documentation of all texts and presentations mentioning historical individuals from the ordinary part of the population. It is discussed how to extend the national population register to an international register.

The NHPR opens new possibilities in research in many different disciplines. Some of these require new methods such as handling privacy when the register includes both the present and the historical population and a flexible ID-system that handles changes in the linking of records in the sources. The paper discusses fundamental research problems such as representativity and the ratio of individuals who are either registered twice or missing from the 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>
- Holden, L. & Boudko, S. (2018). The Norwegian Historic Population Register and Migration, *J. of Migration History* 4, 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 med eksempler fra migrasjon og kildekritikk, Heimen. December 2025
- Holden, L., Boudko, S. & Thorvaldsen, G. (2020). Lenking og kobling i Historisk befolkningsregister, Heimen, 3, 216-229. <https://doi.org/10.18261/issn1894-3195-2020-03-04>

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>

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>

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, ISSN 1044-579X.
<https://doi.org/10.1016/j.semancer.2021.12.001>

Ruggles, S., Cleveland, L., Lovaton, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D., Lee, J., & Merrill, N. (2025). Integrated Public Use Microdata Series, International: Version 7.6 Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D020.V7.7>

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>

Solla T. (2004) National databases and rheumatology research I: longitudinal databases in Scandinavia, *Rheumatic Disease Clinics*, 30 (4), 851-867.

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>