

# The Development of Microhistorical Databases in Norway. A Historiography

By Gunnar Thorvaldsen and Lars Holden

To cite this article: 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>

## HISTORICAL LIFE COURSE STUDIES

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

VOLUME 13, SPECIAL ISSUE 5

### GUEST EDITORS

George Alter  
Kees Mandemakers  
Hélène Vézina



## MISSION STATEMENT

# HISTORICAL LIFE COURSE STUDIES

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

### Methodological Articles

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

### Research articles

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

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

### Co-Editors-In-Chief:

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

### Associate Editor:

Eva van der Heijden (Utrecht University)

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

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



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



HISTORICAL LIFE COURSE STUDIES  
VOLUME 13 (2023), published 11-05-2023

# The Development of Microhistorical Databases in Norway

## A Historiography

Gunnar Thorvaldsen  
UiT The Arctic University of Norway

Lars Holden  
The Norwegian Computing Center

### ABSTRACT

Norwegian work on microdata started out with the full count 1801 census and census and vital records from around the capital. Today, most census and ministerial records from 1801 until the mid-20th century have been scanned, transcriptions are being completed, much is encoded and made available via the websites of the Digital National Archives and UiT The Arctic University of Norway. This article complements a previous publication on empirical results from historical microdata. It is primarily organized by technical issues: digitization of source materials, encoding and standardization, building of the Historical Population Register for the period since 1800, record linkage and source criticism as well as GIS. Presently, partner institutions are building the Historical Population Register with prolonged support from the Norwegian Research Council. This will contain longitudinal records of the nine million persons who lived in Norway since 1800. The register increasingly makes it possible to follow the entire population. Unique personal IDs with corresponding URLs to the person page providing links to many sources introduce a new level of historical documentation. Cross-sectional and vital records are being interlinked with automatic and manual record linkage software. Longitudinal data is available for searching as timelines and in Intermediate Data Structure format from UiT The Arctic University and for searching at Histreg.no, which also caters for manual editing. We are well on the way to creating a database that can fill the void in the two centuries before the Central Population Register starts in 1964.

**Keywords:** Norway, Microdata, Censuses, Church records, Data processing, Population register, Record linkage

e-ISSN: 2352-6343  
DOI article: <https://doi.org/10.51964/hlcs14315>

© 2023, Thorvaldsen, Holden

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

## 1 INTRODUCTION

This article summarizes the history of the computerization of historical microdata in Norway. It is a technically oriented follow up to the historiography article highlighting the research accomplishments using such datasets originating in nominative historical sources (Sommerseth & Thorvaldsen, 2022). These are primarily full count nominative censuses, vital registers in church books and emigration records. The aim is to cover the main developments of national significance, including illustrative regional ones.

The article is primarily organized by technical issues: digitization of source materials, encoding and standardisation, building of the Historical Population Register for the period since 1800, record linkage as well as mapping and administrative borders. We start, however, with a brief overview of the organisation of the databases. Since many users, especially the less quantitatively oriented, are not so familiar with information technology we try to write without too many technicalities. On the other hand, some such details are necessary for the article to be useful for those who are building similar databases.

## 2 ORGANIZING THE HISTORICAL MICRO-DATA INFRASTRUCTURE

### 2.1 EARLY PROJECTS AT THE UNIVERSITIES OF BERGEN AND OSLO

The first historical Norwegian computer project was undertaken at the University of Bergen, aiming to computerize the full count 1801 census of Norway. Started in 1968, this was a joint venture between the Historical Institute, the National Archives and Statistics Norway. The major goal was easier access to one of the most central sources from "the old society." Then Norway was still largely rural with 800,000 inhabitants in the countryside, 80,000 town dwellers and 8,000 civil servants (Statistics Norway, 1980). Rather than converting the source material into numerical codes, it was wisely decided to enter the contents of the census verbatim. This source version soon became popular among archivists and genealogists, as well as social historians. The main guideline was to change as little as possible when transcribing the records onto rolls of punched paper tapes for the Univac mainframe. Student and soon researcher Jan Oldervoll not only supervised the punching but also developed software to print, sort and encode each census column.

A user-friendly Internet interface pioneered the 1801 census onto the World Wide Web and allowed open access to search for place names, personal names, and other census variables — unfortunately, this census contains no information on birthplace. Users were enabled to produce univariate and bivariate statistics, including the analysis of households classified according to the Hammel-Laslett system (Sommerseth, 2011; Statistics Norway, 1980). This 1801 census edition has been employed in a number of local history studies and was the central source for several master theses written in Bergen. They linked the census records interactively to baptisms, marriages and burials from the church books for 48 parishes spread across Norway. The most notable results pertain to the background and destiny of unwed mothers and social mortality differentials (Engelsen, 1983; Haavet, 1982). The Historical Institute at the University of Bergen also computerized parts of other censuses, emigration records, and ministerial records in cooperation with the Regional Archive in Bergen.

The microhistory project *Norwegian Social Development 1860 to 1900* was launched by the University of Oslo in 1971. Microhistory meant both history on the individual level and within a limited framework of time and space. Information about occupation, age and other variables must be available for each person on an individual level, not just aggregated for all residents of e.g., a municipality. In this way it would become possible to aggregate data to the desired level of analysis: families, the inhabitants of a census district, the entire municipality or larger regions. In turn, the life courses of groups of individuals were combined into collective biographies. The studies were thus both cross-sectional and longitudinal. The two selected locations were Ullensaker and the capital Kristiania (Langholm, 1974, 1976). These were the methodological and geographical contexts for studies of workers at the new Kværner Brug factory in Oslo and the radical popular movement started by Marcus Thrane in 1849. Mainframe computers were used both to organize the source material to find data about the individuals in order to describe featured groups statistically and to describe the whole population in the study areas.

The Ullensaker and Kristiania project developed software for transcribing the sources, proofreading, correcting errors, listing and sorting the material. This internally constructed program package HISO was also used to encode the datasets, while the standard statistical program package DDPP (Discrete Data

Program Package) of the mainframe DEC computer was used for aggregates. The computerized sources were the 1865, 1875 and 1900 censuses, parts of the emigrant records, as well as the church registers 1845 to 1875 for Ullensaker parish and the 1875 census for Kristiania which was the biggest component (78,000 individuals) anyway. The large costs of transferring the sources to the computer, could be justified with extensive research use by projects with many graduate students and collective supervision. More than 20 master theses were written based on these digitalizations. In addition, a number of students were inspired to study related themes for other locations. Central among the themes were social and geographical mobility, suitable themes based on the sources' information about occupation and birthplace, utilized with modified or full-fledged reconstitution methodology. Our knowledge increased especially about who moved to America or to the city, about connections between father's and son's occupations and about typical professional careers in the second half of the 19th century. The digitized source materials are stored in the National Archives and in the University of Tromsø.

## 2.2 THE NORWEGIAN HISTORICAL DATA CENTRE

The computerization of historical sources was transferred to the startup Norwegian Historical Data Centre (NHDC) at the UiT The Arctic University of Norway in Tromsø from 1978 onwards. Since 1985 it is a permanent body within the Faculty of Social Sciences and Humanities, serving researchers, teachers, students, and genealogists nationwide. The prime aim of the NHDC and its partners is a national population registry of the 18th and 19th centuries (cf. Section 4). The NHDC published printed books with verbatim transcriptions of the 1865, 1875, 1900 and 1910 census manuscripts as well as parish registers with alphabetical indexes. Also, digital versions following a national standard for data entry and data distribution were sent to the users (Nygaard, 1995). The encoded versions of the censuses were initially distributed on diskettes. In cooperation with the National Archives, the full count census transcriptions were expanded into nationwide datasets.

The transcription of parish registers was a more labor-intensive undertaking than the censuses, making present geographic coverage more limited, at 2/3, but higher in the 19th century. The handwriting in these sources is often Gothic, increasing the transcription difficulties significantly. The source material was made available as originals or xeroxed copies from the National or Regional Archives but now mainly as scanned versions via the Internet.

## 2.3 THE DIGITAL ARCHIVE

Through its Digital Archive, the National Archives present open access documents in digital formats. Many of the digitized sources contain nominative microdata in scanned or transcribed searchable formats, especially from censuses and church books, but also from emigration protocols, probate records and sailors' or military registers. The transcribed copies originate from the UiT The Arctic University of Norway, the National Archives' own transcription groups, international genealogical companies and volunteers. Scanned or transcribed sources can be chosen by period, topic, archive deposit or location of origin. When searching the transcribed versions, individuals can also be retrieved by names, gender, birthplace and -year/-date, status, type of event or role, event time and information about a relative (cf. the web interface page in Figure 1). Users may choose to search for exactly spelled names or for variants.<sup>1</sup> Certified researchers may upon application, program access to the records in the Digital Archive via an API gateway.

Registered users who log in to the Digital Archive are supposed to submit a correction notice if they believe, after inspecting the original, that a transcription is not correct. Users are warned that they may not suggest changes to what is written verbatim in the source. If users wish to enter additional information about a person, or to link information about the same person from several sources, they should use the online interface to the Historical Population Register ([Histreg.no](http://Histreg.no); see Section 4.2). The publication of transcribed church books through cooperation with the genealogical companies Ancestry, My Heritage and Family Search has resulted in a large increase in correction notices. Direct links between transcriptions and scanned images increasingly become available. Another development is that scanned printed and handwritten texts will progressively become searchable by employing OCR and handwriting recognition techniques.

<sup>1</sup> There is a version of the user interfaces in English at [digitalarkivet.no/en](http://digitalarkivet.no/en), but some of the online help information is in Norwegian only.

Figure 1 Search page to retrieve individuals in the Digital Archive

The search page is titled "Search page to retrieve individuals in the Digital Archive". It features a light blue background and a dark blue search bar at the top right with a magnifying glass icon and the word "SEARCH". To the left of the search bar is a "Reset" button. The page is divided into two main sections: "Individual information" and "Event information".

**Individual information:**

- First name:** Input field containing "Ida Mathilde".
- Last name:** Input field containing "Hansd\*".
- Gender:** Input field with a dropdown menu showing "Female".
- Role:** Empty input field.
- Birth year:** Two input fields, the first containing "1886" and the second containing "1886".
- Birth date (mm-dd):** Input field containing "mm-dd".
- Place of birth:** Empty input field.

**Event information:**

- Event year:** Two input fields labeled "From" and "To", each with a dropdown arrow.
- Event date (mm-dd):** Input field containing "mm-dd".

**Related person:**

- First name:** Empty input field.
- Last name:** Empty input field.
- Birth year:** Empty input field.
- Role:** Empty input field.

When browsing scanned archive material in the Digital Archive, users will come across contents that is fully or partly restricted for use via the Internet. This applies to sensitive personal data about racial or ethnic origin, political opinion, religion, philosophical belief, trade union membership, criminal convictions or offenses. In addition, social security-like numbers can be blocked. However, place of birth, date of birth, citizenship, marital status, occupation, place of residence and place of employment are ordinarily not considered personal matters. In child welfare and adoption cases, the duty of confidentiality only expires after 100 years, and information in the state census manuscripts is obtained for use for statistical purposes only for 100 years according to the Statistics Act. The National Archives assume that the Personal Data Act does not provide protection for the deceased, and that sensitive personal data, can ordinarily be made available on the Internet when all the persons referred to have certainly died. Because it is difficult to establish that all persons referred to are dead, in many cases the 100-year rule applies anyway.

## 3 TRANSCRIPTION AND CODING

### 3.1 TRANSCRIPTION

The basic principle when transcribing censuses and other sources has been to copy the content as literally as possible. The chief method for achieving this aim has been to proofread the transcriptions, letting one assistant read from the digital transcription while another checks against the original source. We found this to be superior to double transcription because some transcribers tend to reproduce errors. And there are ergonomic reasons: proofreading is a less strenuous exercise than typing. As a last step we have an "acceptance control", where essential information (name, age) in 10% of the records are proofread to check that the proportion of errors is acceptable. Basically, this is a method to allow experienced transcribers to correct mistakes that were missed by less experienced colleagues during proofreading. Also, we sort the material to spot variable values with frequency one which indicate rare, erroneous cases and we program the computer to look for illogical combinations of variable values such as teenage widows. See the record linkage section 6.3 for techniques to spot conflicting information when we combine data from several sources.

However, there are exceptions to the verbatim transcription rule. We introduce a distinction between first names and surnames and between occupations and household positions. Census takers often used a special sign to indicate that the information is the same as from the previous person, which is replaced by the information itself. Unreadable handwriting is marked with double question marks, and illogical information with double exclamation marks. When in doubt about two different transcription alternatives, these can be separated by the masterspace "@". There are cases where the information is conflicting inside one census record, for instance when a "daughter" is marked as male in the gender field. Based on names and other information, the transcriber should correct what is obviously an error in the source and flag the correction in the comments field. Such checks are also built into the transcription apps which nowadays are constructed with standard database packages such as MS Access, rather than the "home brew" which was popular in past decades. To sum up: transcription rules are a compromise between creating a verbatim copy and enhancing the user-friendliness of the resulting database (Thorvaldsen et al., 2015).

## 3.2 HISTORICAL OCR PROJECTS

When starting to transcribe censuses and church records at the Norwegian Historical Data Centre in 1978 we aimed for a decentralized and low-tech solution: typing the content of the sources with ordinary typewriters. The typewriters were equipped with OCR-B balls, resulting in a special font that the rudimentary OCR readers at the time could recognize. This kind of OCR acted as a missing link after the heyday of the Hollerith punch card and before PCs became common. Anyone interested in this aspect of the history of computing should read *Travels in Computerland* (Schneider, 1974). We mailed our typed pages to a commercial company in Sweden who returned the contents as ASCII text files, for checking and correcting on a mainframe computer. The advent of PCs for transcription soon made this OCR-B setup with external companies obsolete. The advent of general OCR program packages for the PC later on made OCR topical again. The affordable software Omnipage let us convert printed text with diverse fonts to machine readable files. The Norwegian Institute of Local History has compiled a list of printed historical sources containing some 2,000 entries, useful when selecting material for optical character recognition. A few pilot projects used OCR to computerize printed source copies, but nowadays it is more efficient to let local historians and other volunteers transcribe source materials with their PCs. A noted example is from the National Archives' collection of medieval documents, with the 19th century publication *Diplomatarium Norvegicum*. Indexes to the collection were constructed by transferring the printed versions to text files with OCR. This increased the typical number of documents used for a dissertation from a few to a hundred. Another significant collection of digital source material only mentioned here even if it may contain microdata, is the collection of computerized records retrieved from public agencies by the National Archives and preserved in secure systems there (Thorvaldsen, 1992).

## 3.3 STANDARDIZING AND ENCODING THE CENSUSES

The transcribed censuses exist in a verbatim full-text version as well as in an encoded standardized format. It contains numeric codes for occupation, family status, and parish or municipality of birth, and no numeric codes were transcribed from the sources. In first instance, the main purpose of the coding of the censuses was to create statistics. There are several reasons for not simply using the aggregates published after each census. Boundaries between the administrative census units as well as the categorizing of occupations and other information between the censuses often changed, making historians' comparisons over time difficult. A third reason is that the variables in published statistics are combined at group levels, not at the individual level, increasing the risk of introducing ecological fallacies, i.e. jumping to conclusions based on aggregates (Langholm, 1976). A fourth reason is that the standardized codes make it easier to link people from source to source, for example code 0724 for the birthplace is more consistent over time than changing municipality names for the same locality.

While simple fields such as gender, marital status and age require little standardization and few rules, coding occupations, family status, places of birth and ethnicity are complex tasks (Thorvaldsen, 1994). After verbatim transcription, we use computer programs to semi-automate the coding. By eliminating identical versions of the same occupation etc., thus compressing the source entries of each variable into frequency lists, the coding became more consistent and less time-consuming. Each person entry is equipped with relevant codes, creating a standardized version of the census that can be used on its own in a statistics program or together with the verbatim text version of the source for record linkage.

Relationships between people in the same household or family are coded with specially constructed variables by the IPUMS project after they received our abovementioned encoded versions. For example, a relationship variable provides information about the spouse because it reciprocally contains the ID numbers of the spouses of married persons in each household. Using these ID numbers makes it clear which husbands and wives belong together, even if there were several couples in a household. Corresponding variables "point" from the children to each of the parents. However, in large, complicated households, the IPUMS computer program that creates these variables may introduce errors, which can be corrected at the [Histreg.no](http://Histreg.no) website. The relationships between household members in the census lists allowed us to distinguish between the 19th century decline in the number of farm servants, and the domestic servants whose numbers did not decline until World War II (Thorvaldsen, 2008).

"Place of birth" is an important variable both because it can distinguish between people with common names, and it helps us to get an overview of the life course of migrants. A Norwegian census will usually specify a person's place of birth at the municipality or parish level. All municipalities in the same county have the first two digits of the four-digit code in common, which simplifies the study of migration. When the third digit is zero, it means that the relevant municipality is urban. Thus, historically there were never more than 9 towns or 99 municipalities in a province. To track administrative border changes, the numbering system is dynamic. Each area that has ever been a separate municipality has been assigned a unique code. There are lots of municipality changes, while the counties seldom changed. To handle immigrants, the system has been expanded with country codes for the rest of the world.<sup>2</sup> Problems can arise with ambiguous names of municipalities and other place names, as we shall see below. Sometimes, the name of a territory that includes several municipalities is indicated. If the entire area lies within the same county, the county code can be used.

Table 1 presents an overview of the encoded data from the censuses in 1801, 1865, 1875, 1900 and 1910 and hopefully soon 1920 as they are available via the internet from our partner at the University of Minnesota ([ipums.org](http://ipums.org) and [nappdata.org](http://nappdata.org)).

Table 1 *Coded text strings and population figures in the national censuses 1865 to 1910 and Sandefjord town in 1920. The full count resident population in 1920 was 2,649,775*

Census variable	1865	1875	1900	1910	1920
Family/household	19,188	22,000	25,172	20,050	1,390
Occupations	75,734	78,000	368,598	334,079	2,044
Birthplaces	41,813	55,000	73,893	58,227	1,083
Population	1,701,756	1,813,424	2,240,092	2,391,782	5,764

Standardization of name strings is useful for quantitative analyses of name frequencies as well as for linking data records. Both first and last names have been standardized, in collaboration with professor of Nordic languages Gulbrand Alhaug (2011), in a research council-funded project. This happens via a model with three levels:

A. Graphemic level <i>Orthographic variants</i>	B. Phonemic level <i>Linguistic variants</i>	C. Lexicographic level <i>Standardized name</i>
Caroles Carolus Charolus	Karoles Karolus	Karolus

The project prepared a list of personal names where variant spellings and nearby linguistic variants were standardized to the same standardized name. For example, Kristian with K and Ch was coded to the name Kristian, and Fredrek is standardized to Fredrik, while the difference between Anne and Anna was preserved since the variants are pronounced differently and both variants are common. Since the standardization is rather conservative, it is necessary to use an additional program that calculates the phonetic distance between name forms, using the Jaro-Winkler algorithm (Winkler, 1990). Thus, personal records with similar name forms will be linked even though the initial standardization was conservative. A comparison of the effect of name standardization on linking in the US and Norwegian censuses showed that the number of links in the Norwegian ones increased by 17% due to the standardization and that it was the work with surnames that contributed the most (Vick & Huynh, 2011).

2 For Norwegian and foreign historical place codes, cf. [https://rhd.uit.no/koding/fs\\_koder.html](https://rhd.uit.no/koding/fs_koder.html).

Another factor is that naming customs change over time (Fure, 1990). This affects both the frequency of different names in different parishes, and which forms of name are perceived as synonymous.

"Farm data" found in the 1838 and 1886 tax lists were transcribed with OCR or manually and are available on the Internet. The census of 1865 and 1875 also include agricultural data. Compared with the tax lists they are more complete, including cottars *and* farmers and also provide information on the seeding and the number of animals. However, the reliability of these figures is weakened because the informants feared taxation. So, one may assume that the census data on sowing and animal husbandry give a deflated picture. But it is still realistic to construct a measure for relative production at the farms and cottars' places. In principle, we considered two calculation methods, either the monetary value of the production or its nutritional content. Because there was no real market for potatoes, grain etc. in much of Norway, we based the calculations on the nutritional value of which the number of calories was the most important aspect (Statistics Norway, 1880). These were calculated and considered in detail in the research to assess the value of the introduction of potatoes in Norwegian agriculture (Lunden, 1975).

For husbandry we use the calculation method from Statistics Norway for the 1875 census. The conversion of the relative value of different livestock into cow-entities is based on sales value for adult animals and thus takes account of local differences (Thorvaldsen, 1995b, pp. 486–488). Again, we use Lunden's (1975) calculations to estimate the farms' and cottars places' production value as thousands of calories per year. In the 1865 census, the agricultural figures were entered in the same form as the rest of the information and are thus linked to individuals, mostly heads of household. This is more complicated in the 1875 census because agricultural data was noted on a separate form but can be linked automatically to persons and places. For instance, a calculation of agricultural output was performed in the community history of Kvenangen municipality to assess differentiations in production connected to the three ethnic groups in the area, Sami, Fins and Norwegians (Bjørklund, 1985).

## 4 THE HISTORICAL POPULATION REGISTER (HPR)

### 4.1 THE REGISTER

The HPR is becoming a national register covering the 9.5 million people, who lived in Norway sometimes during the period 1801–1964 with an estimated 87 million person records in the most important cross-sectional, vital events and migration sources (Holden, Boudko, & Thorvaldsen, 2020). The population register has an open access part including only deceased persons in open sources. Except for the deceased from 1928 to 2014, there are few open sources after 1920. The closed part contains restricted sources and persons still living. Both parts are linked to the Central Population Register (CPR) started in 1964. Automatic record linkage has been made at the UiT The Arctic University of Norway and at the Norwegian Computing Center (NR). NR has also developed the website [histreg.no](http://histreg.no), where manual links in the open part are added by volunteers. The links between the open and the closed parts are not public, however. Most censuses and church records from 1800 to 1960 have been transcribed and the open parts are becoming available in the Digital Archive (Holden et al., 2020; Thorvaldsen, 2011b).

The motivation for building the HPR is to provide a central national infrastructure for research in history, social sciences, medicine and a number of other disciplines. The HPR also has an important cultural component of cooperation about the tracing of genealogies for over 200 years, and by comparing inconsistent records it fulfills a crucial source critical goal. In Section 6 we will present methods by which automatic links are created between instances of the same person in different sources and pointers to relationships between family members. This is done both by UiT and NR. In addition to the above-mentioned institutions, the National Archives (the Digital Archive), the Institute of Public Health (digitizing and linking 20th century sources), Statistics Norway (access to the Central Population Register), the National Library (support of and interaction with local historians) and the Norwegian School of Economics (tax records) are project partners. The Historical Population Register recently received its second significant funding from the infrastructure program of the Norwegian Research Council.

Linking of instances of persons in various sources has traditionally been carried out in the context of farm and family history for rural municipalities with detailed studies of the local church registers and censuses supplemented with other written and oral sources. The studies have mapped settled farmers to a greater extent than people without real estate and people who moved. In most cases,

such work is carried out by people with detailed insight into the local conditions. Traditionally, the work has been carried out manually, but computers have increasingly been used to streamline and systematize the work with the source material and the analyses. Machine representation also simplifies the communication of the results (Kjelland, 2018), usually in community history books. It has long been a goal among professional historians to activate the community history books' detailed farm and family genealogies in historical research (Hovland, 1977), and we believe that the cooperation with the National Library will promote this aim.

## 4.2 OPEN ACCESS TO THE HISTORICAL POPULATION REGISTER: HISTREG.NO

The open part of the Norwegian Historical Population register is available at the website [histreg.no](https://histreg.no). Introduced in 2016, the National Archive is responsible for the site, which is developed by the Norwegian Computing Center. Histreg.no may be considered as an index to the Digital Archive of the National Archive, with transcriptions of church books, censuses and emigrant lists with about 57 million person-records from the period 1800–1920, as well as sources like prison and health records, school protocols, etc. Histreg.no has a page for each person entry in all these sources. If several person entries are considered to belong to the same person after record linkage, the pages are merged so that the person page presents the life course, including a list of all sources belonging to the person. Each person gets a unique ID generated from the unique source entry ID provided by the National Archive. The use of this unique ID in scientific articles and elsewhere enhances the documentation of the data used in research. The URL to the person page of the explorer Fridtjof Nansen is <https://histreg.no/index.php/person/pf01073681015788> where the last 16 digits are his unique ID which in this case originates from the 1920 census in the Digital Archive. For each source entry, there is a hyperlink to the transcribed source in the Digital Archive and the scanned source image.

Figure 2 shows a typical person page in histreg.no, the top showing gender, name and information about the birth and death. This information may be edited by a logged in contributor. Then follows links to the person pages of parents, siblings, partners and children with information about family relations from the sources and additional family relations added manually. The life course table lists the linked source records about the person with hyperlinks to the transcribed sources in the Digital Archive. This information may only be changed by adding or removing source records, not by changing the data in the sources. At the bottom of the page (not shown in Figure 2), it is possible to write a brief biography, explain the linking, specify references or add other comments about the person.

The record linkage algorithms to create unique persons from several appearances in the sources are developed by the UiT The Arctic University of Norway and the Norwegian Computing Center, where the record linkage rate varies by source. It is close to 90% between the 1910 and 1920 censuses where we have families and birthdates and significantly lower for sources with less information. The algorithms are based on comparing the names, birthdate or -year, birthplace, address and family relations (see Section 6). In addition, volunteers make links and family relations manually in a crowd sourcing effort. During recent years, more than 170 persons made about 40,000 links per month corresponding to the output of two persons working full time in the same period. Still, more than 90% of links are made by algorithms. By March 2023, the Histreg database contains 12,2 million links between source entries for 3,7 million persons (i.e., persons found in at least two sources). We estimate there were 6,6 million persons living in Norway in the period 1800–1920 (Statistics Norway, 1995). Each manual link and family relation has a time stamp and identification of the contributor. The same person is sometimes registered with conflicting data in the sources. The system, therefore, includes the option to register a link as verified in spite of conflicting information, in order to block the link from accidental removal during quality controls. Histreg also lists conflicting data that are not verified, which manual contributors are encouraged to check manually. We expect to continue adding millions of links during the following years, but Histreg will never be "complete".

Histreg.no has many features to improve the quality of the dataset and encourage its use. It is possible to refer to persons that are not yet identified in the sources. The program may list the largest unlinked families in a specific census for a municipality in order to encourage further linking in a region. In addition to the sources including the whole population we include some thematic registers, such as war prisoners 1940–1945. There are also links to the Local History Wikipedia and biographic data in newspapers.<sup>3</sup> During 2023, we shall add information on wartime sailors and Norwegian politicians

3 [https://lokalhistoriewiki.no/wiki/Lokalhistoriewiki:Hovedside/Om\\_Lokalhistoriewiki](https://lokalhistoriewiki.no/wiki/Lokalhistoriewiki:Hovedside/Om_Lokalhistoriewiki) and <https://www.retrievgroup.com/about-us>

from 1814 onwards provided the legal restrictions are heeded. Complementary sources provide further information about the persons, enhance the value of each thematic register and increase the interest in Histreg since mutual links make the thematic register more visible. At the same time, the protection of privacy must be respected for persons still alive.

Persons linked in Histreg are not representative of the entire population, as in all historical population registers, since linkage rates are higher for persons with relatively high social status, a permanent address or (slightly) being male. However, the representativeness of statistical results from Histreg can be increased based on data from the full count censuses. Histreg is both an editing and a retrieval tool. In addition to personal information fields such as name and age, the user can refine the search based on type of event, role and year or period and geography — municipality of birth or residence. It is possible to show what partners and parents were related to the retrieved persons. Histreg.no can sort the search results by first name, surname and year of birth. The search also shows the number of interlinked person records.

Figure 2 An editable person page in Histreg with keywords in English

## Ida Mathilde Mikkelsen Hansdater

♀ Kvinne  
 Født: 08.07.1886, Sem sogn Born  
 Død: 06.04.1962, Ukjent sted Dead

[Rediger persondata](#) [Søk etter lignende personer](#) [Edit person page](#) [Search for similar persons](#)

**Foreldre og søsken** Parents and siblings

Far: Hans Jörgen Kristensen Christensen , 1853 -  
 Mor: Anette Andrine Regina Kristensen Kristensen\* ChristensDater , 1861 -  
 Søsken: Kristian Hansen Hansen  
 Søsken: Agnes Hansen  
 Søsken: Anna Marie Hansdtr.  
 Søsken: Hans Ragnvald Hanssen , 1888 -  
 Søsken: Johan Arne?? Hanssen , 1893 -  
 Søsken: Wilhelm Kristensen\* Wilhelmssen , 1894 - 12--  
 Søsken: Hilda Anette Hansen , 1896 -  
 Søsken: Georg Kristensen\* , 1896 -

**Partner og barn** Partner and children

Partner: Gjert Mikael Mikkelsen , 1884 -  
 Barn: Helmer Thorvaldsen , 1912 -  
 Barn: Gulborg Thorvaldsen , 1914 -  
 Barn: Maud Ida Thorvaldsen , 1918 -

[Vis lenking av familiemedlemmer](#) [Show linking of family members](#)

Life course table - Overview of sources with this person

### Livsløpstabell - Oversikt over kilder med personen

Nr	Dato	Kilde	PFID	Rolle	Navn	Fødselsdato	Fødested	Bosted	Fam. stil.	Sivst.	Yrke
1	01.01.1891	FOLK, Sem, 1891-1891	pf01052816001475 Census		Ida Mathilde Hansdater	1886	Sem	Sem: Langerød (Husmondspl)	Datter		
2	03.12.1900	FOLK, Sandar, 1900-1900	pf01037148003376 Census		Ida Matilde Hansen	1886	Sem	Sandar herred: Lystad	d	ug	Datter
3	01.12.1910	FOLK, Sandefjord, 1910-1910	pf01036489005150 Census		Ida Mathilde Mikkelsen	08.07.1886	Sem sogn	Sandefjord: Kongens gate 37	hm	g	Hustru
4	01.12.1920	FOLK, Sandefjord, 1920-1920	pf01073803003878 Census		Ida Mathilde Thorvaldsen	08.07.1886	Langerød i Sem Vest	Sandefjord: Torvgaten	hu	g	Husmor
5	06.04.1962	DODR, 0000, 1951-2014	pc00000001724177 Church record	avdød	Ida Mathilde Thorvaldsen	08.07.1886					

## 4.3 OTHER USER INTERFACES

### 4.3.1 LINKED PAIRS OF CENSUS RECORDS

The simplest way to follow people over time is to link two points in the life course, for example a baptism and a census or two censuses. The latter is done for the censuses of 1865, 1875, 1900, 1910 and 1920. The first three of these enumerations were linked and made available by the Minnesota Population Center as part of the North Atlantic Population Project (NAPP). The website [ipums.org](http://ipums.org) contains data files for Norway with links between the censuses 1865 and 1875, 1865 and 1900 as well as 1875 and 1900, which have been imported into the Historical Population Register. In addition, NAPP includes linked records combining the complete US 1880 census with seven US census samples as well as Norwegian census records (cf. [https://international.ipums.org/international/linked\\_data.shtml](https://international.ipums.org/international/linked_data.shtml)). The Norwegian and American linked censuses have been used together in research on the economy of emigration (Abramitzky, Boustan, & Eriksson, 2012, 2013). On [nappdata.org](http://nappdata.org), the linked censuses can be downloaded in a simple data format i.e., using one record for two linked data records; this simple data structure is possible when only two points in time are covered in each life cycle.

In order to avoid constructing erroneous biographies by linking records that actually belonged to different people, a conservative linking strategy was chosen by the NAPP project, which resulted in low linking rates (Ruggles, Fitch, & Roberts, 2018). The linking strategy for the Norwegian censuses 1865–1875, 1875–1900 and 1865–1900 depends on four time-invariant variables: year of birth, four-digit municipality birthplace code, standardized first name and standardized surname. Birth years were allowed to differ by up to three years for linking men, and up to five years for married couples. Some municipal border changes were neutralized by including the neighboring municipality. To avoid creating a biased selection by under-prioritizing the linking of singles, information on family members was not used – except when linking married couples. Unfortunately, single women were more difficult to link, due to the lack of birthdates in these censuses (Thorvaldsen, 2011b). If the same data record was linked to two different records in the 1875 or 1900 censuses, because these censuses combined a de facto and a de jure count of both resident and present population), information on permanent residents was preferred (Thorvaldsen, 2006). The early Norwegian immigrants to the US, coming before the keeping of systematic migration protocols, have been listed. We have also successfully traced emigrants to Sweden and to north-western Russia (Naeseth & Hedberg, 1993–2008; Thorvaldsen, 2011a; Thorvaldsen & Erikstad, 2007).

### 4.3.2 TIMELINES TO FOLLOW INDIVIDUALS AND FAMILIES

The UiT The Arctic University of Norway presents a system to display "timelines" with longitudinal information from the Historical Population Register (Thorvaldsen, Sommerseth & Holden, 2020). The 1865, 1875, 1900, 1910 and soon the 1920 censuses are available for search on the UiT website via a simple and an advanced user interface, see <http://rhd.uit.no>. After finding a person in one of the enumerations, clicking the house symbol displays information about the household. Linked individuals are equipped with a marker (🏠), shown in the left margin in Figure 3, meaning that further information is available from other sources via unique source references generated by the National Archives.

By clicking the link marker, an overview of the linked data records will be displayed from other censuses and church records. Users are warned that the links are generated automatically, and it cannot be ruled out that the software has introduced erroneous links and that some are missing. For example, Thorvald Mikkelsen was easily identified in the censuses in both 1865, 1900 and 1910, but was not *automatically* linked to his entry in the 1875 census because no close relative was present to be used as linkage criteria. Since he remained on the same farm, the linking was easily done manually. He was adopted as a foster child by the childless farmer who bought the farm from his parents. Clicking on the + sign in the column on the left in Figure 4 displays information about the entire household to which the person belonged in the relevant census year.

The timeline function makes it less time-consuming to follow groups of people over time. Cohorts of people who share the same characteristics can be defined in the advanced user interface, so that the search results are more adapted to statistical purposes. However, there are no built-in statistical procedures, and the user must create the categories herself. An example of a more complex timeline, which also contains information from the church records about fisherman Haldor Hansen (1850–1922) can be found at <https://rhd.uit.no/folketellinger/tidslinje.aspx?idi=8231280>.

Figure 3 *Family on the farm Sjuvestok in Stokke parish in 1865 with linkage symbols in the left margin*

Name	Family status	Marital status	Occupation	Birth year	Place of birth
 Mikkel Nilsen	hf	g	Gaardbr. og Selveier	1812	Stokke Pr.
 Elen L. Hansdatter	Hans Kone	g		1823	Stokke Pr.
 Hans Mikkelsen	Deres Søn	ug	Matros	1844	Stokke Pr.
 Mathias Mikkelsen	Deres Søn	ug	Hjelper Fdr. med Grbr.	1847	Stokke Pr.
 Thorvald Mikkelsen	Deres Søn	ug		1849	Stokke Pr.
 Rikard Mikkelsen	Deres Søn			1852	Stokke Pr.

Figure 4 *Timeline with references to four censuses for Thorvald Mikkelsen. It can be expanded for each census at <https://rhd.uit.no/folketellinger/tidslinje.aspx?idi=3177832>*

	1865	FT1865	0720-001-0027-00-005
	1875	FT1875	0720-001-0103-00-003
	1900	FT1900	0718-002-0075-00-001
	1910	FT1910	0706-005-0038-05-001

For advanced statistical purposes, the advice is to use the Intermediate Data Structure (IDS), which was specified by researchers and data providers who need to transfer records between collaborators (Alter, 2021; Quaranta, 2021). For qualitative purposes the Linked Pair approach or the Time Lines described above, provide a simpler introduction to using the HPR. The IDS has been successfully implemented for regional data from Northern Norway at the Norwegian Historical Data Centre and used to find ground-breaking results on intergenerational infant mortality, in an international project also studying regions in Sweden, Belgium and the Netherlands with comparable datasets (Quaranta & Sommerseth, 2018; Sommerseth, 2018). This is not the place to describe the qualities of the IDS system, which is well documented elsewhere (Alter & Mandemakers, 2014). However, IDS is not a standard tool in Norwegian research on historical microdata. The development of alternative models for data exchange between the partners may serve the internal project needs and may also be a way to distribute microdata to researchers in Norway. But this will certainly not function as well as IDS to promote internationally comparative research projects. IDS is designed for family reconstitution data and has less advantage for census data. As the Historical Population Register now adds relatively more information from vital registers, IDS will likely prove more useful.

## 5 GIS AND MAPPING

### 5.1 OVERVIEW

The starting point for work with GIS in connection with microdata in Norway was the Municipality Database (*Kommunedatabasen*) built and maintained by the Norwegian Centre for Research Data containing statistical information about the municipalities since the 18th century.<sup>4</sup> This database was exploited successfully by the Princeton Project-related fertility decline study, at the time when full count microdata for Norway was still not complete (Sogner, Fure, & Randsborg, 1984; Sogner, Randsborg, Fure, & Walloe, 1986). Attached to the database is a dynamic collection of national maps showing the province and municipality borders dynamically over time since their creation in 1837, which can be ordered and downloaded for a number of GIS software platforms. The UiT extended the municipality map backwards to 1801 with the parish boundary maps from the transcribed 1801 census (Statistics Norway, 1980). Besides, the National Archives, the Institute of Local History and others

4 <https://www.nsd.no/finn-data/kommunedatabasen/>

have cooperated to develop a dynamic cadaster of farms and their placement in the administrative boundary structure since the 18th century, but this project has not yet come to fruition.

A considerable part of Norwegian historical research deals with local areas, municipalities, provinces and other regions below the national level. Therefore, it is important to pose some questions regarding geographical delimitations: How shall the area of study be defined in relation to the surrounding area? What criteria might be used to partition the selected area internally into smaller zones for more detailed study? These questions are closely related to the use of GIS software to display characteristics of population differentials on choropleth maps. The divisions into geographic entities were made through a combination of political, administrative and juridical decisions. More recently infrastructure, social and cultural divisions play a role as well, although language and ethnicity are seldom used as border criteria in Norway.

Both because of and in spite of the rather drastic alterations in municipality boundaries, the four volume *Tromsø City History* is based on its present-day borders. Pedagogically, most readers are more familiar with current municipality boundaries and the financing of the project would have been difficult if the historical work covers areas outside the present municipality or excludes part of it. In volume II Astri Andresen (1994) used the census microdata from 1801, 1865, 1875 and 1900 to reconstruct the present-day boundaries of the municipality for the long 19th century. She selected the data for the specific farms or places situated in the area that today is a part of Tromsø. These contiguous pieces then became censuses covering the present-day Tromsø municipality over time. In contrast to the printed aggregates from Statistics Norway, where the old boundaries were followed, she made overviews of parts of the area or all of it. Thus, we get to know how many people lived in today's Tromsø-area on past census days, the birthplaces of the inhabitants according to 19th century censuses, etc. The ethnic composition of the population was studied by analysing the Sami areas in peripheral parts of the municipality. There are valid reasons why one should use the present-day administrative divisions by back-projecting them to earlier periods: Access to the sources might be easier, the provenance principle decreeing that source material is to be organised by the present-day administrative divisions. Thus, a discussion of the choice of region is imperative, but is often lacking in commissioned research.

The reduction in the number of Norwegian municipalities by nearly 50% of their maximum number, confirms that boundaries on a low level are not stable over time. The boundaries of the provinces were traditionally more stable than the borders of the municipalities, but they too have been subject to small changes, sometimes when the municipality-boundaries have been changed. However, this changed dramatically when a conservative Parliament in 2017 supported a regional merger reform act that affected 13 out of 20 provinces, an unpopular measure in most provinces. After new elections, the centrist parties dominated Parliament and voted to dissolve most of the mergers. Future historians will, therefore, face extra hassles when mapping developments on the provincial level around 2020.

The complicated process of determining the borders between administrative units started in medieval times and will not be detailed here. It must suffice to start with the Laws of Local Democracy (*Formannskapslovene*) of 1837, which mainly based the new municipalities on the old ecclesiastical divisions into parishes and sub parishes (*prestegjeld* and *sogn*). At the time, the border surveys were provisional due to lack of resources to stake the frontiers with locals in the field or examine the relevant archives. The established boundaries only gradually became more certain, as borders were regularly revised based on new information. A game changer was the detailed descriptions of the provinces in *The Country and People of Norway* (Helland, 1876–1917).

## 5.2 SPLITTING MUNICIPALITIES FOR ANALYSIS

In connection with local micro-studies and as a basis for comparison, it is necessary to divide the area of study into lesser units. Here the commissioned researcher will have more freedom. When writing his volume III of the history of Oslo, *The Divided City*, Knut Kjeldstadli (1987) discusses several ways of dividing Norway's capital, singling out five factors that to varying degrees have contributed to the actual formation of the city's districts: boundaries, names, local institutions, administrative divisions and social conditions.

A source-oriented approach may work to some degree for rural municipalities. In the *History of Balsfjord and Malangen Municipality* Hauglid's (1981) attempt to characterize the social history of each particular census ward on the basis of the machine-readable versions of the nominative censuses of 1865 and 1900. In the chapter called "Ethnic Diversity in a Multi-Cultural Society" the census of 1865 is the main source along with Friis' ethnographic maps from 1861.<sup>5</sup> The author examined each of

5 <https://www.dokpro.uio.no/friiskartene/1861/1861oversikt.html>

the four census tracts to map settlement patterns, the distribution of Norwegians, Sami and ethnic Fins and their kinship relations. Aggregative figures were directly derived from the census, although with explanations based on other source materials. Because of changed boundaries this 1860s snapshot is not comparable with the one from 1900 with 16 census wards. This unfortunately urged the author not to aim for consistency and comparability over time, but rather highlighted occupations than ethnicity. Thus, even though we get much insight into local ethnicity as well as trades and industry, a systematic description of industrial and ethnic developments is regrettably lacking.

In *The History of Sandefjord* Finn Olstad (1995) took this one step further by attempting to divide Sandar municipality, which used to surround that town, into a coastal and a land-locked part. His purpose was to investigate any differences in trades and industries in separate parts of the municipality according to the 1900 census. About half were employed in the primary sector, mainly agriculture in the land-locked part of the municipality, while this was only true for about a fourth of the population in the coastal part, rather partaking more in maritime trades. However, this was based on the census wards which often do not follow such a divide, but rather extend from the coast into the inland areas. An alternative strategy has been to rather use the railway line and the main road as demarcation lines parallel with the coast, but this necessitates the use of farms and other place entities in the census to divide the municipality into a coastal, an internal and a middle zone between the railway and the main road. Such a definition is easy to relate to for most readers (Thorvaldsen, 1997).

With its Ward Database (*Kretsdatabanken*) Statistics Norway and The Norwegian Centre for Research Data made information on the level of census wards more accessible for researchers, but only for the post-war censuses. The Wards Database mirrors the official statistics from the censuses on the municipality level: the population's gender and age structure, industry and employment, religion and language, as well as housing with up to 500 variables. Thus, the census wards which only had a practical function for the census takers, became statistical units of analysis. As part of this, Statistics Norway redesigned the boundaries of the census wards in order to create meaningful entities, e.g., by distinguishing between urban and rural areas. This could render aggregates from the wards less comparable over time unless researchers re-aggregated the information according to the new boundaries. For foreign researchers census ward data is an interesting alternative, since the Nordic countries do not participate in the IPUMS.org project with microdata from the post-war period. However, the online database [microdata.no](http://microdata.no) where users can create anonymized aggregates, is not available to researchers without a Norwegian social security id-number (Ballo, 2019). The relative inaccessibility of the latter dataset does not warrant a detailed description here.

### 5.3 CHRONOLOGICAL ADJUSTMENT OF THE MUNICIPAL AGGREGATES OR BOUNDARIES

In national analyses, social scientists and historians have often employed the municipality as a unit for geographical data. This is because the municipal level is seen as less random and more pedagogical than other divisions, and large amounts of statistics and other information are available on the municipal level. However, the constant changing of municipal borders must be solved if we wish to use these data sets for comparison over time. A standard solution is integrated in the Municipality Database of the Norwegian Centre for Research Data, containing variables about Norwegian municipalities from 1769 onwards. When extracting a time series, researchers can request that the municipal boundaries are to be standardized to a chosen year. The software will then "move" a proportion of the population affected by the border changes in the period studied. This method can be illustrated with a simple example from election studies, where we want to study the development of voting from 1933 to 1936. This is complicated because an area with 100 persons was transferred from the municipality Fjord to Fjell in 1935. If we do not have data for the transferred area, we must assume that the votes from the group of 100 were distributed in the same way as the votes in the entire municipality. If party A received 70% and party H received 30% in Fjord in 1933, the program moves 70 A-votes and 30 H-votes from Fjord to Fjell in 1933 before comparing with the 1936 results, obviously a gross approximation with potentially misleading results. Often, the transferred areas are peripheral rather than central parts of the municipality and have a different employment structure. The splitting of a municipality into two new ones can be similarly difficult to approximate, whereas the merging of such administrative areas is more straightforward.

Ideally, we should study migration and other social phenomena in relation to the smallest locations, that is, to the farm or the building (Thorvaldsen, 1995a), which is difficult, since the census only reports place of birth on the municipal level. The study of migration in Troms province from 1865 to 1900 was complicated

by the changes in the parish and municipality boundaries. Thus, an important prerequisite when studying migration, is to create a consistent division of the province into municipalities over time. The census of 1865 was taken for thirteen parishes in Troms province, while later censuses used a more fine-grained division. For this reason the rough parish division of 1865 was used as a basic structure to make comparisons over time. By "moving" the farms and other places into the 1865 parish structure, it was possible to merge the smaller municipalities in later censuses into a common structure for the whole period.

When we transfer people from one municipality to another to compensate for changes in the municipality borders, we must be certain to change people's place of birth accordingly.

## 5.4 THE MAP PORTALS URBGIS AND BERGIS

The UrbGIS map portal shows historical maps for the urban areas in Norway and is integrated with the 1900 and 1910 censuses in the Digital Archive. The map for Bergen in UrbGIS mainly covers the period 1881–1957 based on street names and numbers. The richer collection of historical maps for the city of Bergen in BerGIS covers the period approximately 1830–1881 using both the address system and a cadastral system. The map portals can also show map layers for applicable property boundaries (see <http://urbgis.uib.no/> and <http://urbgis.uib.no/bergis/>).

## 5.5 ENUMERATING THOSE ABSENT AND VISITING

In the censuses from 1875 onwards, some persons according to the instructions were enumerated twice. These were people travelling or away from home for other reasons, most often to work in a different location or to live preliminarily away from home. This has consequences for our statistical use of the census microdata. Since no one should be counted twice, we must choose between the two entries. And a similar decision must be made during record linkage as discussed in Section 6.

Earlier, in the 1865 full count census, census takers were instructed to enumerate residents "where they sleep" and not include "Anyone residing temporarily in a place, ...". Thus, this was a *de jure* census, unlike the British at the time, which noted people where they happened to be on census night, i.e. *de facto*. Problems, especially with the registration of sailors, may explain why combined *de jure* and *de facto* enumeration was introduced into the 1875 census. Inspired by the international statistical conferences, the 1875 census forms contained a special field for visitors in order to note the "usual residence of those, who on the 31st of December temporarily stayed overnight in the house." For those temporarily absent, a special section was included at the form's bottom, an arrangement dropped in later censuses. Statistics Norway detailed the definitions of temporary residence or absence, e.g., about absent lodgers. Students and servants in 1875 should be enumerated *de jure* where they studied or worked. In later censuses, however, the students were considered permanent residents in their municipality of origin, temporarily absent from there and temporarily present where they studied. As their numbers increased after 1960, the university cities successfully lobbied for a change in the enumeration rules in order to receive state funding according to population size. Thus, for the 2001 census, statistics Norway again enumerated students where they lived when studying (Thorvaldsen, 2006).

When analysing the censuses from 1875 onwards with microdata, it is important to choose whether to exclude either the persons absent or those visiting. Including both leads to over-enumeration because of duplicates. Including those absent results in the resident or *de jure* population, while including those visiting gives the present or *de facto* population. In-between groups such as the abovementioned students can be fitted into either category according to national specifications. Especially in municipalities with many sailors or fishermen, the difference in population size according to the definition of the resident and the present populations can be significant, even exceeding a tenth of the population. It follows that men predominated (see Table 2). It is possible to perform record linkage between records about those visiting and those absent in the same census. Complete matching of the groups cannot be expected, since many of those absent had left Norway and there are foreigners among those visiting. Also, inaccuracies produce the usual problems linking nominative microdata. If we demand record linkage matches where the years of birth and birthplace codes are identical, and a Jaro-Winkler similarity between first names in the records for the visiting and absent groups, these are matched in 29,618 cases. If we demand exact match on birthdate, the number decreases to 20,210 — it is natural that there are more problems with birthdates for these mobile groups than among those who were enumerated at home. Further analysis of these records should be undertaken in order to assess the proportion of inconsistencies between enumerations by different census takers at the same point in time.

Table 2 *Number of persons by gender and residential status in the 1910 census for Norway*

<b>de facto/de jure</b>	<b>Sum</b>	<b>Female</b>	<b>Male</b>
At home	2,256,281	1,205,084	1,051,183
Visiting	133,591	29,708	103,881
Absent	82,222	25,902	56,317
Sum	2,472,094	1,260,694	1,211,381

## 6 RECORD LINKAGE SUMMARY

Many articles and several books have been written about the linking of historical individual data. We covered basic points above in connection with the Historical Population Register. It is an impossible task to discuss all the rules and experiences here, but based on our practice the main principles can be summarized in ten points (Fure, 2000; Thorvaldsen et al., 2015).

### 6.1 TEN COMMANDMENTS OF LINKING HISTORICAL PERSONAL RECORDS

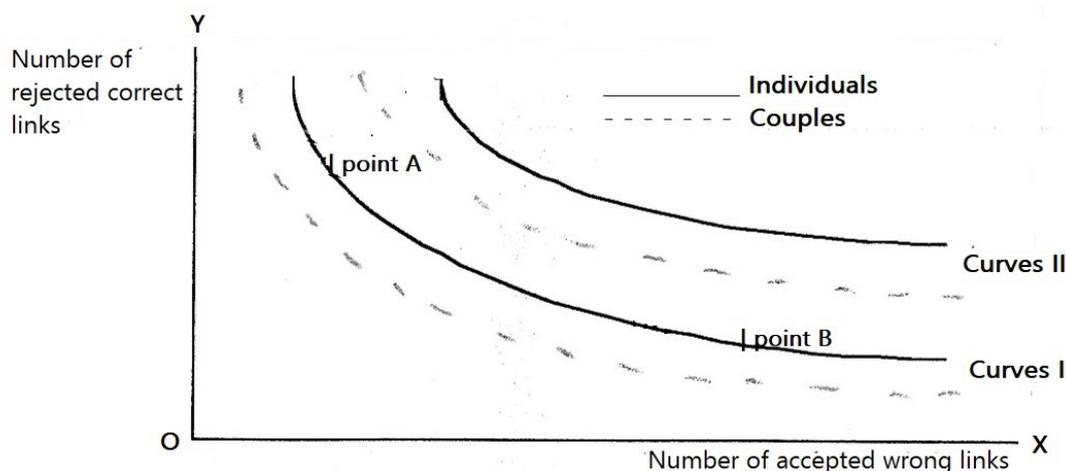
1. The life course that the links describe must be logical. Obviously, a marriage record must precede a burial record.
2. Historical, individual level microdata contains too many inconsistencies to demand full agreement between the variables used for linking records from two or more sources. For example, there will often be variant spellings of names or places of birth and errors in the year of birth.
3. Only accept reasonable discrepancies between the source entries based on experience, e.g., with name variants. Kristine and Kristina should be considered true name variants, but Anna and Anne are both frequent and distinct.
4. We cannot link duplicate identities, only classify them as linking candidates. An example is several persons named Ole Olsen born in Oslo in 1851.
5. Search for complementary source material in case of doubt (Point 3 and 4). For example, birthdates in the 1910 and 1920 censuses could provide a basis for linking back to a baptism list with birthdates that could not be linked to the 1900 census.
6. Variables that vary throughout the life course may be used source critically for linking, the most relevant time-variant information being addresses and occupations. We can therefore link an apprentice carpenter's record to a later entry with a master carpenter even if there are other competing entries with different occupations but otherwise duplicate information.
7. Relationships to other people may also be time-variant, but can still be used critically for linking, even if information on group relations is not constant over time in the sources. Links made with time-variant variables should be flagged (cf. the end of Section 4.2).
8. The links can be changed manually, usually based on information in genealogists' records. However, this must be flagged and documented, and it may have consequences for other links in the database (cf. Point 10). Links made according to Rules 6, 7, 8 may create bias in the linked sample of records.
9. The precision of protocol data generally improved over time. Thus, you can place greater emphasis on small differences in names, year of birth, etc. when linking records with similar information around the year 1900 than around the year 1800 and still avoid the risk of linking duplicates (cf. Figure 5).
10. Some links will be broken because new sources are added, algorithms are improved or genealogies checked. It can potentially lead to many changes — like a nuclear proliferation in the database but will usually have smaller consequences. Even so, the consistency of the links in the database must be periodically checked with detailed algorithms.

Below we shall deal with typical examples of problems and solutions that arise when linking the 1910 and 1920 censuses into the Historical Population Register. These censuses are the first that contain exact dates of birth. This provides unique opportunities for analysing discrepancies between different sources with personal data. When linking retrospectively, the 1900 and 1891 censuses as well as the church records can be brought in to resolve discrepancies. In this way we fulfil parts of the source-critical purpose, which together with the more empirical one, are the main motives for creating a historical population register.

Manual linking provides flexibility and makes it possible to consider special cases that are only described in the tradition of each individual family, local historical knowledge or special name forms of the nickname type. In addition to being more efficient, automatic record linkage using computer programs will ensure more uniform handling of the sources, but this can more easily introduce errors due to lack of consistency and uniqueness in the sources. Experience shows that the algorithms must be detailed in order to provide a constant proportion of correct links with time and place due to variations in the sources such as degree of name similarity, migration and other factors. When record linkage is based on the information in two or three source variables, these data items must be relatively unique and consistent with respect to name, date of birth, year of birth and place of birth as is often the case in such recent sources as the censuses from 1910 and 1920, and where we can check against supplementary sources and data on related persons.

The linking techniques are a compromise between creating as many correct links as possible while not introducing erroneous links, which is well illustrated for Denmark in Figure 5, a country with source material similar to Norway (Johansen, 2002). Here the relationship between the number of false positive (i.e., incorrect) links on the x-axis and the number of false negative (i.e., missing) links on the y-axis is illustrated. The main idea in Figure 5 is that stricter linking criteria will limit the number of false positives and increase the number of false negative links. Looser rules will have the opposite effects, increasing the number of false positive problem links, while limiting the number of false negatives. Johansen maintained that the increasingly exact content of the source material made it easier to find a compromise between the two considerations than was the case in earlier sources, illustrated by the two different sets of curves, marked I for the later and II for the earlier periods. The introduction of birth dates in the 1910 and later censuses was another step towards origin (O) in the diagram.

Figure 5 Linking individuals and couples



Note: Curves I: later periods, Curves II: early periods. Point A: Strict linkage rules, Point B: looser rules. Adopted from Johansen (2002).

## 6.2 RECORD LINKAGE PROBLEMS WITH THE 1910 AND 1920 CENSUSES

As stated above, the censuses of 1910 and 1920 were the first to ask for the date of birth for the entire population. The background for this addition was that from 1903 population registers were started in the municipalities, needing more precise identifiers. Name information in the censuses had become more unstable since marrying women changed their surname, and many persons abandoned the custom of patronymics and changed to other types of family names, for example the name of their farm. In addition, the number of persons born in cities increased rapidly during this period. The 1910 census

was transcribed in collaboration between the UiT The Arctic University of Norway and the National Archives. The combination of professional transcription staff and sources with good readability should guarantee a high quality of the work. The 1920 census was mainly transcribed by volunteers who were "paid" with insight into this source before the 100 years exclusion period expired on 3 December 2020. The relevant forms were scanned and posted in the Digital Archive with access for approved users.<sup>6</sup> The quality of the scan is good, but the colour differences are gone, which makes it harder to distinguish between what is written during fieldwork and what has been added by local administrators or Statistics Norway. Altogether, there is reason to expect a somewhat lower transcription quality compared with 1910. It also plays a role that the 1920 census consists of one-person forms, where it is more difficult to decipher writing by comparing individual entries than when we have forms listing many persons, like in 1910. This can especially have an effect in town censuses because these were not filled in by rural teachers, but usually by the house owners themselves. However, we must bear in mind that a systematic overview of discrepancies in the 1900 and earlier census versions shows that most inconsistencies are not a result of transcription errors but of original source differences (Fure, 2000).

### 6.3 SURNAMENES, PLACES AND DATES OF BIRTH PROBLEMS IN THE 1920 AND 1910 CENSUSES FOR SANDEFJORD TOWN

#### 6.3.1 DATE OF BIRTH

Of the 1,101 persons that were linked for Sandefjord between the 1910 and 1920 census based on name, place of birth code and year of birth, 180 had discrepancies in the date of birth. Although the "Date of birth" is a good distinguisher and helpful to find errors and inconsistencies in other information this variable is not always reliable either. We can further compare with the 1900 census which gave the date of birth for persons up to two years old. Trygve Gjertsen was born in 1900 in neighbouring Larvik town. His date of birth was the 1st of April in the 1920 census, but the 7th of April in both the 1910 and the 1900 census. A check against the 1920 census scanned image shows that a more likely transcription is the 7th of April. Eivind Halvorsen (PID pf01036489007059)<sup>7</sup> born in 1900 in Sandefjord had two conflicting birthdates: 5th of June, 6th of June and again 6th of June in the three censuses 1900–1920. Since a question mark was added to the date in the first of these censuses, that is likely the wrong information. For Emil Larsen, born 1899 in Tjølling (PID pf01036489003252), the discrepancy is once more due to a transcription problem. As can be seen from Figure 6, the day digit '8' looks like '5'. Incidentally, Emil was called Hartvigsen after father Hartvig in the 1900 census, but in 1910 the whole family had changed their surname to Larsen.

Figure 6 Example of unclear birthday in the 1920 census for Emil Larsen

1 desember 1920.

Skjema 1. Personseddel nr. 3

Sandefjord by. Tellingskrets nr. 8

Husliste nr. 238 Husholdningsliste nr. 2

1. Fullt navn: Emil Larsen

2. Mannkjønn<sup>1</sup>. Kvinnekjønn<sup>1</sup>.

3. Fødselsdag 28-5 i året 1899

4. Fødested: Tjølling

Opgi herred eller by i Norge eller fødeland utenfor Norge.

6 <https://media.digitalarkivet.no/en/ft/browse?censuses%5B%5D=18&counties%5B%5D=07&municipalities%5B%5D=0706&text=>. The 1910 census is scanned and available in the Digital Archive, useful for control purposes.

7 We have added PIDs so that readers can study the examples by inserting them in Nansens url (cf. Section 4.2).

### 6.3.2 SURNAMENES

By 1920 it had become the norm to have a different surname than by birth, especially for women upon marriage. Ida Mathilde (PID pf01052816001475), born in Sem in 1886 experienced even four different surnames according to the censuses from 1891 to 1920. She was first entered with the traditional female patronymic Hansdatter in 1891, then with the male and more modern version Hansen in 1900. After getting married in 1910 the name became Mikkelsen, changed to Thorvaldsen in the 1920 census. Family tradition tells that the last change was not due to remarriage, but that the husband competed with another transporter named Mikkelsen, and therefore changed his surname to a patronymic based on his father Thorvald Mikkelsen. Divorces were rare, but due to remarriages, Ida Mathilde was far from unique. By linking the women using first names, date of birth and place of birth code, we can see how unusual it was to keep the surname. In Sandefjord in 1920, we found that only five out of 58 married women had the same surname as they had in 1910, and this was because they married men already holding the same surname.

### 6.3.3 PLACE OF BIRTH

The "Place of birth" was usually given at the municipality level – the parish had been abandoned as a census unit in the late 19th century but was still used in the church records. Spelling inconsistencies make standardization necessary, so all census records are equipped with a four-digit municipality code.<sup>8</sup> Ole Kristian Kristoffersen (PID pf01036504009121) born 1848 in Sandar municipality was easy to link from the 1920 to the 1910 census using a consistent name and date of birth, but the place of birth according to the first-mentioned source created doubts, he was allegedly born in Veøy in Nordmøre, north of Bergen. Figure 7 shows that this was not a straightforward transcription error, rather, Statistics Norway's standardization of the information interpreted the farm name "Westad" as belonging on the west coast of Norway ("Veøy N Møre") and did not take into account that the same farm name exists in Sandar municipality, surrounding Sandefjord. Thus, a proximity principle needs to be applied.

A typical misinterpretation on the part of Statistics Norway concerns Kvelle, which is routinely placed in neighbouring Hedrum municipality with its sub-parish of that name, while earlier censuses placed Augusta Larsen as born in 1857 in Kville, Sweden. In contrast, there is an interesting but uncertain linkage candidate in the 1875 census for Aker bordering on Oslo, where there is a Gustav Adolf Olsen (PID pf01073803007686) born 1864 in Fredrikshald. When searching the Digital Archive, a more reliable candidate appears: Sausage maker Gustav Alfred Olsen born 1871 in "Vorter Aker" had fathered a child in Sandefjord in 1920. "Vorter" is not Norwegian, and inspection of the scanned edition of the church book shows "Vestre (Western) Aker", but the handwriting is indistinct, making it difficult to decipher this birthplace for the transcribers from India or China.<sup>9</sup> The church records are often scanned from microfilm copies, which results in poorer legibility than the direct scanning of the original forms.

Figure 7 Statistics Norway violating the proximity principle by defining the farm name to be located in the western part of Norway

1. Fullt navn: Ole Christian Christoffersen

2. Mannkjønn<sup>1</sup>. Kvinnekjønn<sup>1</sup>.

3. Fødselsdag 15-7 i året 1848

4. Fødested: Westad Veøy Nordmøre } Opgi herred eller by i Norge eller fødeland utenfor Norge.

8 This coding of the 1920 census is done by Gunnar Thorvaldsen for the town of Sandefjord, as the first municipality in Norway in preparation for this article.

9 Birth register for Sandefjord parish 1916–1932 (0706Q). The born and baptized 1920, page number 63, line number 40.

## 7 CONCLUSION

In half a century, work on microdata in Norway has grown from activity in two universities transcribing and researching the full count 1801 census and a collection of census and vital records microdata from around the capital. Today, most census and ministerial records from 1801 until the mid-20th century have been scanned, transcribed, coded and made available via the websites of the National Archives and the UiT The Arctic University of Norway. Encoded and interlinked census records are also available from Minnesota Population Center as part of [ipums.org](http://ipums.org). Many master and doctoral theses as well as research articles have been written on topics within social history and historical demography. Presently, research and administrative partner institutions are building the Historical Population Register with prolonged support from the Norwegian Research Council. This will contain longitudinal records of the nine million persons who lived in Norway since 1800. The register will make it possible to follow the entire population in Norway for up to seven generations where we previously could only follow samples in a few municipalities for a shorter period. Unique personal IDs with corresponding URLs to the person page provide links to many sources and introduce a superior level of historical documentation. Cross-sectional and vital records are being interlinked with automatic and manual record linkage software. Longitudinal data is available for searching as timelines and in Intermediate Data Structure format from UiT and for searching at [Histreg.no](http://Histreg.no), which also caters for interactive editing. Much record linkage and quality assurance work remain but we are well on the way to creating a database that can fill the void in the two centuries before the Central Population Register starts in 1964.

## ACKNOWLEDGEMENTS

We thank the editors of the special issues of *Historical Life Course Studies* for their comments and good advice during our writing of this article.

## REFERENCES

- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review*, 102(5), 1832–1856. doi: [10.1257/aer.102.5.1832](https://doi.org/10.1257/aer.102.5.1832)
- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration. *Journal of Development Economics*, 102(C), 2–14. doi: [10.1016/j.jdeveco.2012.08.004](https://doi.org/10.1016/j.jdeveco.2012.08.004)
- Alhaug, G. (2011). *10 001 navn: Norsk fornavnleksikon* [10 001 names: Norwegian encyclopedia of first names]. Oslo: Cappelen Damm.
- Alter, G. (2021). Reflections on the Intermediate Data Structure (IDS). *Historical Life Course Studies*, 10, 71–75. doi: [10.51964/hlcs9570](https://doi.org/10.51964/hlcs9570)
- Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal historical microdata, version 4. *Historical Life Course Studies*, 1, 1–26. doi: [10.51964/hlcs9290](https://doi.org/10.51964/hlcs9290)
- Andresen, A. (1994). *Tromsø gjennom 10000 år, Bind 2: Handelsfolk og fiskerbønder 1794–1900* [Tromsø during 10,000 years, Volume 2: Merchants and fishing farmers 1794–1900]. Tromsø: Tromsø commune.
- Ballo, J. G. (2019). Microdata.no: New technology allows instant access to Norwegian register data. *Tidsskrift for samfunnsforskning*, 60(4), 398–408. doi: [10.18261/issn.1504-291X-2019-04-04](https://doi.org/10.18261/issn.1504-291X-2019-04-04)
- Bjørklund, I. (1985). *Fjordfolket i Kvæningen: Fra Samisk samfunn til Norsk utkant, 1550–1980* [The fjord people in Kvæningen: From Sami society to Norwegian outskirt, 1550–1980]. Tromsø: Universitetsforlaget.
- Brosveet, J., Olaussen, T. G., & Sande, T. (1979). *Kommuneendringer 1838–1978* [Municipality changes 1838–1978]. Bergen: NSD rapporter.
- Engelsen, R. (1983). Mortalitätsdebatten og sosiale skilnader i mortalitet [The mortality debate and social differences in mortality]. *Historisk Tidsskrift*, 62(2), 161–202.

- Fure, E. (1990). Personnavn og tidsånd [Personal names and the spirit of the times]. *Namn og Nemne*, 7, 35–55.
- Fure, E. (2000). Interactive record linkage. The cumulative construction of life courses. *Demographic Research*, 3, 1–20. doi: [10.4054/DemRes.2000.3.11](https://doi.org/10.4054/DemRes.2000.3.11)
- Fure, E. (2004). ... *En besynderlig regelmæssighed: Dødeligheten i Asker og Bærum på 1800-tallet med særlig vekt på spedbarnsdødeligheten* [... A curious regularity: Mortality in Asker and Bærum in the 19th century with particular emphasis on infant mortality] (Doctoral dissertation). Oslo: University of Oslo.
- Haavet, I. E. (1982). *Avvik eller uhell? Ugifte foreldre omkring 1800 — En sosial analyse*. [Deviation or accident? Unmarried parents around 1800 — A social analysis] (Master thesis). Bergen: University of Bergen.
- Hauglid, A. O. (1981). *Balsfjorden og Malangens historie* [The history of Balsfjord and Malangen]. Storsteinnes: Balsfjord kommune.
- Helland, A. (1876–1917). *Norges land og folk: Topografisk-statistisk beskrevet* [Norway's land and people: Topographically-statistically described]. Kristiania: Aschehoug.
- Holden, L., Boudko, S., & Thorvaldsen, G. (2020). Lenking og kobling i Historisk Befolkningsregister [Record linkage and family pointers in the Norwegian Historical Population Register]. *Heimen*, 57(3), 216–229. doi: [10.18261/issn1894-3195-2020-03-04](https://doi.org/10.18261/issn1894-3195-2020-03-04)
- Hovland, E. (1977). *Folket, bygda og historia* [Population, community and history]. Bergen: Universitetsforlaget.
- Johansen, H. C. (2002). Identifying people in the Danish past. In H. Sandvik, K. Telste, & G. Thorvaldsen (Eds.), *Pathways of the past. Essays in honour of Sølvi Sogner on her 70th anniversary 15 March 2002* (pp. 103–110). Oslo: University of Oslo.
- Kjeldstadli, K. (1987). Hva er en bydel [What is a city part?]. *St. Hallvard*, 2, 5–15.
- Kjelland, A. (2018). Mapping and analysing remigration based upon Norwegian farm- and genealogical history projects. *Journal of Migration History*, 4(2), 314–329. doi: [10.1163/23519924-00402005](https://doi.org/10.1163/23519924-00402005)
- Langholm, S. (1974). Historie på individnivå. Omkring Ullensaker-undersøkelsen — Et mikrohistorisk eksperiment [History at the individual level. About the Ullensaker investigation — A microhistorical experiment]. *Historisk Tidsskrift*, 53, 243–272.
- Langholm, S. (1976). On the scope of micro-history. *Scandinavian Journal of History*, 1(1–4), 3–24. doi: [10.1080/03468757608578894](https://doi.org/10.1080/03468757608578894)
- Lunden, K. (1975). Potetdyrkinga og den raskare folketalsvokst i Noreg frå 1815 [Potato cultivation and the faster population growth in Norway from 1815]. *Historisk Tidsskrift*, 54, 275–313.
- Naeseth, G. B., & Hedberg, B. (1993–2008). *Norwegian immigrants to the United States: A biographical directory, 1825–1850* (Vols. 1–5). Madison, WI: Vesterheim Genealogical Center and Naeseth Library, c1993-.
- Nygaard, L. (1995). *Histform — Norsk standard for registrering og utveksling av nominative folketellingsdata for årene 1865-1910* [Histform — Norwegian standard for transcription and exchange of nominative census data for the years 1865-1910]. Tromsø: Registreringsentral for historiske data, Universitetet i Tromsø.
- Olstad, F. (1995). *Sandefjords historie. Strandsitter og verdensborger* [The history of Sandefjord. Coastal resident and citizen of the world] (Vol. B. 1). Sandefjord: Sandefjord kommune.
- Quaranta, L. (2021). Reflections on the use of the Intermediate Data Structure (IDS) in historical demographic research. *Historical Life Course Studies*, 10, 76–80. doi: [10.51964/hlcs9571](https://doi.org/10.51964/hlcs9571)
- Quaranta, L., & Sommerseth, H. (2018). Introduction: Intergenerational transmissions of infant mortality using the Intermediate Data Structure (IDS). *Historical Life Course Studies*, 7, 1–10. doi: [10.51964/hlcs9288](https://doi.org/10.51964/hlcs9288)
- Rokkan, S., & Valen, H. (1966). Archives for statistical studies of within-nation differences. In S. Rokkan (Ed.), *Data archives for the social sciences* (pp. 112–127). Paris: Mouton.
- Ruggles, S., Fitch, C. A., & Roberts, E. (2018). Historical census record linkage. *Annual Review of Sociology*, 44, 19–37. doi: [10.1146/annurev-soc-073117-041447](https://doi.org/10.1146/annurev-soc-073117-041447)
- Schneider, B. R. (1974). *Travels in Computerland: Or, incompatibilities and interfaces: A full and true account of the implementation of the London Stage Information Bank*. Reading, MA: Addison-Wesley Publishing Company.
- Sogner, S., Fure, E., & Randsborg, H. B. (1984). *Fra stua full til tobarnskull: Om nedgangen i barnetall i norske familier i de siste 200 år, med særlig vekt på perioden 1890-1930* [From a housefull to parity two: On the decline in child numbers during the last 200 years, emphasizing 1890–1930]. Oslo: Universitetsforlaget.

- Sogner, S., Randsborg, H. B., Fure, E, & Walloe, L. (1986). Le déclin de la fécondité en Norvège (1890–1930) [The decline in fertility in Norway (1890–1930)]. *Annales de Démographie Historique*, 1986, 361–375. doi: [10.3406/adh.1987.1669](https://doi.org/10.3406/adh.1987.1669)
- Sommerseth, H. (2011). *Northern co-residence across generations. In northernmost Norway during the last part of the nineteenth century* (Ph.D. thesis). Tromsø: UiT The Arctic University of Norway. Retrieved from <https://munin.uit.no/bitstream/handle/10037/3372/thesis.pdf?sequence=5&isAllowed=y>
- Sommerseth, H. (2018). The intergenerational transfer of infant mortality in northern Norway during the 19th and early 20th centuries. *Historical Life Course Studies*, 7, 69–87. doi: [10.51964/hlcs9284](https://doi.org/10.51964/hlcs9284)
- Sommerseth, H., & Thorvaldsen, G. (2022). The impact of microdata in Norwegian historiography 1970 to 2020. *Historical Life Course Studies*, 12, 18–41. doi: [10.51964/hlcs11675](https://doi.org/10.51964/hlcs11675)
- Statistics Norway. (1880). *Statistik angaaende det Norske jordbrug. Fornemmelig i femaarsperioden 1871–1875 og i året 1875* [Statistics on the Norwegian agriculture. Especially 1871–1875]. Kristiania: Trykt i Ringvolds bogtrykkeri.
- Statistics Norway. (1980). *Folketeljinga 1801. Ny bearbejding* [Population census 1801. New processing]. Oslo: Statistisk sentralbyrå.
- Statistics Norway. (1995). *Historisk statistikk 1994* [Historical statistics 1994]. Oslo: Statistisk sentralbyrå.
- Thorvaldsen, G. (1992). The preservation of computer readable records in the Nordic countries. *History and Computing*, 4(3), 211–227.
- Thorvaldsen, G. (1994). The encoding of highly structured historical sources. *Computers and the Humanities*, 28, 301–305. doi: [10.1007/BF01830278](https://doi.org/10.1007/BF01830278)
- Thorvaldsen, G. (1995a). Longitudinal sources and longitudinal methods — Studying migration at the Stockholm Historical Database. In A. Brändström & L.-G. Tedebrand (Eds.), *Swedish urban demography during industrialization* (pp. 219–250). Umeå: Umeå University.
- Thorvaldsen, G. (1995b). *Migrasjon i Troms i annen halvdel av 1800-tallet. En kvantitativ analyse av folketellingene 1865, 1875 og 1900* [Migration in Troms province in the second half of the 19th century. A quantitative analysis of censuses 1865, 1875 and 1900] (Doctoral dissertation). Universitetet i Tromsø, Tromsø. Retrieved from <https://hdl.handle.net/10037/1390>
- Thorvaldsen, G. (1996). *Håndbok i registrering og bruk av historiske persondata* [Handbook on transcription and use of historical person data]. Oslo: Tano Aschehoug.
- Thorvaldsen, G. (1997). On boundaries and areas in local historical research. In J. E. Myhre & K. Kjeldstadli (Eds.), *Festskrift til Sivert Langholm*. Oslo: The Norwegian Historical Association. Retrieved from <https://hdl.handle.net/10037/29041> (web version in English)
- Thorvaldsen, G. (2006). Away on census day. Enumerating the temporarily present or absent. *Historical Methods*, 39(2), 82–96. doi: [10.3200/HMTS.39.2.82-96](https://doi.org/10.3200/HMTS.39.2.82-96)
- Thorvaldsen, G. (2008). Hushjelper og jordbrukstjenere — Når kom nedgangen i tjenerallene? [When did the number of domestic servants decline in Norway?]. *Historisk Tidsskrift*, 87(3), 451–464. Retrieved from <https://hdl.handle.net/10037/2014>
- Thorvaldsen, G. (2011a). Household structure in the multi-ethnic Barents region: A local case study. In D. G. Anderson (Ed.), *The 1926/27 Soviet polar census expeditions* (pp. 117–132). Oxford, New York: Berghahn.
- Thorvaldsen, G. (2011b). Using NAPP census data to construct the Historical Population Register for Norway. *Historical Methods*, 44(1), 37–47. doi: [10.1080/01615440.2010.517470](https://doi.org/10.1080/01615440.2010.517470)
- Thorvaldsen, G., & Erikstad, M. (2007). Utvandring til Sverige på 1800-tallet [Emigration to Sweden in the 19th century]. *Heimen*, 44(2), 117–130.
- Thorvaldsen, G., Pujadas-Mora, J., Andersen, T., Eikvil, L., Lladós, J., Fornés, A., & Cabré, A. (2015). A tale of two transcriptions. Machine-assisted transcription of historical sources. *Historical Life Course Studies*, 2, 1–19. doi: [10.51964/hlcs9355](https://doi.org/10.51964/hlcs9355)
- Thorvaldsen, G., Sommerseth, H. L., & Holden, L. (2020). Anvendelser av Norges historiske befolkningsregister [Interfaces to Norway's historical population register]. *Heimen*, 57(3), 230–243. doi: [10.18261/issn.1894-3195-2020-03-05](https://doi.org/10.18261/issn.1894-3195-2020-03-05)
- Vick, R., & Huynh, L. (2011). The effects of standardizing names for record linkage: Evidence from the United States and Norway. *Historical Methods*, 44(1), 15–24. doi: [10.1080/01615440.2010.514849](https://doi.org/10.1080/01615440.2010.514849)
- Winkler, W. E. (1990). String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. *Proceedings of the Section on Survey Research Methods, American Statistical Association* (pp. 1–8). Retrieved from <https://eric.ed.gov/?id=ED325505>