

The Norwegian historic population register and migration

Lars Holden, Svetlana Boudko,

Norsk Regnesentral, pb. 114 Blindern, NO-0314 Oslo, Norway

Corresponding author: Lars.Holden@nr.no

The Norwegian historic population register and migration

Abstract

This paper describes the development of a Norwegian historical population register which is the first open national register. There have been 9.7 million residents of Norway in the period 1735 and 1964 and 37.5 million events in the most relevant sources.¹ We want to link together as many events as possible for the same persons and families but only include links that have a high probability to be correct. The linking is performed by automatic methods and crowdsourcing.

A national population register is particularly important for migration studies. It is necessary to have a national register in order to follow migration inside Norway. There is often a stepwise national migration preceding international migration, making a national register also necessary for describing international migration. In addition, many international sources specify place of birth by country. Therefore, it requires a national register in order to identify the same person in Norwegian sources.

Keywords: National population register, migration, Norway,

Introduction

Several regions are building and extending historical longitudinal population registers for research, statistical and historical purposes. This paper describes the construction of a Norwegian historical population register (HPR). Most of the register is available on the internet. We follow the population of Norway for more than two centuries. Since most sources include location, we will be able to trace the migration. The present population is slightly above 5 millions. Hence, the register will be much larger than the closed Icelandic register,² the regional registers in Sweden^{3,4} and the samples from the Netherlands.⁵ Unlike other registers, HPR will utilize automatic linking of sources, expert linking and crowdsourcing on the internet. Linking is defined as connecting information about the same person or relatives from different sources. Many events about the same person describe the lifecycle of the person.

¹ Thorvaldsen, Gunnar. *Using NAPP Census Data to Construct the Historical Population Register for Norway*, Historical Methods. 44 (1) 2011, s. 37–47.

² deCode, www.decode.com

³ The Demographic Database (DDB), Umeå University <http://www.ddb.umu.se>

⁴ Rotemannsarkivet, persons in Stockholm 1878-1926, <http://www.stadsarkivet.stockholm.se/>

⁵ Historical samples of the Netherlands <https://socialhistory.org/en/hsn/index>

Migration is a central part of the population register. In the period 1820-1930 about 900 000 Norwegians emigrated, one of the highest ratios in Europe.⁶ The Norwegian population was 2 240 000 in 1900.⁷ There has also been a migration inside Norway looking for new farmland and jobs and as a result of urbanisation. The population of the capital Oslo increased from 9 000 in 1800 to 228 000 in 1900. In addition, persons with important positions and their families moved between different regions in their life time and over generations. A national register is necessary in order to describe all these types of migration. The register will follow families some generations after the family left Norway and some generations before the family moved to Norway when we are able to document this in sources. Following families for generations also internationally is genealogical interesting and increases the interest for the register.

It will always be easier to link together events with persons with rare names, special positions and that do not migrate. Utilizing crowdsourcing by users increases the bias. Our ambition is to make as many links as possible where each link has a high probability. This is in contrast to others⁸ where statistical representativity is given more weight at the cost of not using address and family relationship in the linking. When this information is not used, some of the most well documented links are not included. When we make statistical estimates for the population, e.g., number of birth per women, it is necessary to correct for the bias. We believe our approach will give better estimates since our estimates will be based on more links, i.e. a larger data set.

The first version of the population register⁹ was released in November 2015 and by October 2016 there are about 27 mill. person source entries, 1,6 mill. links and 0,7 mill. persons with links in the database. We expect that the number of links will increase slowly before we have recruited many contributors, then faster for several years and then slower again. The number of links depends also on how many new sources we are able to add and the focus on development of automatic linking algorithms. The number of events in potential sources is huge. The register will, of course, never be completed. Our aim is to link 80 per cent of the population (i.e. found in minimum two sources). The current status is about 10 per cent. It is important to be restrictive with new contributions such that the quality improves steadily, instead of degrading due to wrong links. The need for the longitudinal HPR extending over more than two centuries has been documented in a series of research papers from a wide range of scholarly fields: medicine,¹⁰ demography, history, and the social sciences including economics. Research topics range from migration studies, the tracing of genetic diseases via name studies to ethnic differentials with respect to cohabitation.

⁶ The National Archive, Emigrant protocols, <http://www.arkivverket.no/arkivverket/Arkivverket/Statsarkivet-i-Oslo/Norges-dokumentarv-Statsarkivet-i-Oslo/Emigrantprotokoller-fra-Oslo-1867-1966>

⁷ Statistics Norway, <https://www.ssb.no/a/fob2001/utstilling/1900/resultater.html>

⁸ NAPP, www.nappdata.org/napp.

⁹ www.histreg.no

¹⁰ Hovig, Eivind. *Norske gener?* i *Genialt*, nr 1/2010, s. 14–15.

http://www.bion.no/filarkiv/2010/07/Genialt_1_2010_norske_gener.pdf

There are many possible extensions of the HPR database. One of the most important extensions is to include an overview over all houses and farms. This makes it possible to follow persons and families living at the same place during their life time and over generations. More precise documentation of the location of each house, possibly with coordinates, makes it easier to document national migration. This will also improve the quality of the HPR database. We will include as many sources into the database as possible in order to improve the quality and increase the value of the database. We try to make the database as open and transparent as possible focusing on documentation and references to high quality sources. We want it to be important for the scientist, for the experienced genealogist and for the general public. With a unique ID and URL for each historic person, HPR will be a natural reference for texts about the not famous Norwegian historical persons. It is possible to write biographies about each person in HPR, but we encourage to refer to other texts instead.

A previous plan for the HPR database based on wiki-technology has been described earlier.¹¹ Now we use a relational database solution with an application programming interface (API) calls to connect to the Digital Archive of the National Archives that provides updated data. This gives a more effective database administration, easier to handle contributions, but less flexibility in building up new uses of the same data.

The three periods

The register is divided into three parts due to legal reasons. The oldest part of the register is open and new sources are added when these may be disclosed. The Statistics Act requires that the censuses are closed for a period of 100 years. The Personal Data Act similarly protects the church records for 60 years, while sensitive issues such as adoptions are protected for 100 years. Practical issues are also important since the church records are sent to the Regional archives after 80 years. Information from the municipal censuses may be disclosed after 60 years, but little of this material has been digitized. These sources may include information about living persons, but we are not allowed to link information about living persons in the open register due to the Privacy regulations.¹²

The second period is closed and continues until 1964 when the National Population Register starts. The second period may include all sources. Here, there is no restriction on linking except that this may only be performed by professionals since the register is closed. The third period is the National Population Register. This describes all persons with permanent resident in Norway and has been updated continuously since 1964. It includes family relationship and migration.

The register will not contain sensitive information. As an example, we plan to include the death register of Norway from 1928-1960, but leaving out the cause of death. Researchers that need data from the closed periods will get access to this in anonymous form from the Statistics Norway provided the project is accepted after a scientific and legal evaluation. The

¹¹ Holden, Lars; Thorvaldsen, Gunnar; Bråthen, Torkel Rønold. *Historisk befolkningsregister og DNF 1814*. Heimen (ISSN 0017-9841). 49 pp 399-414. 2012.

¹² With, Christian. *Slektsforskning og personvern*. Slett og Data 1, p. 11-13, 2005.

register may be combined with data from other sources, possibly with more sensitive data. The National Archive is responsible for the first two periods while Norsk Regnesentral develops the database for HPR for the first period. The third period is the responsibility of the Norwegian Tax Administration.

Input data

The register includes on as many sources as possible. There are four different methods to include the sources in HPR and the methods are prioritized in the following order:

1. HPR includes data from the Digital Archive and the private depository Digital Inn of the National Archive using API. This implies that update of this data is automatically available in HPR. It includes transcribed national and municipal censuses, church books, emigration and immigration records, probate register, prison register, tax records etc.
2. We also include data that is not available in the Digital Archive and store this in a HPR database with proper references and links to other databases. This may be private registers of high quality registering events that are not documented in the Digital Archive. In addition, sources that may not be disclosed in the open Digital Archive but are included in the second period are of this category.
3. Contributors to HPR are encouraged to register certain types of data in the HPR database. Currently this includes death, marriages, anniversaries, and events of public interest from newspapers. We have registered about 0.6 mill. events using automatic recognition of data from Aftenposten, the largest newspaper in Norway, from the period 1950-1999. Contributors to HPR can register similar events from all newspapers. Norwegian newspapers have a history back to 1860. However, it is more relevant to document events from the last 100 years that are not documented in other open sources. This makes it possible to document persons living in the last period and include this in the open database. It is also possible to register persons directly in an input table provided: (i) the persons are not registered in any of the sources listed above or the sources we expect to include in the near future and (ii) these persons are identified by a trusted contributor like Busetnadssoge.¹³ These persons are listed in a separate person input table sorted by the contributor and then the name of the person.
4. For each person registered in the sources described above, it is possible to make references to any other source. This may include church books that are not transcribed, tax records, internet sites with local history¹⁴ or documenting the ancestors of the persons participating at the first national parliament,¹⁵ international records, family records, local historical books, scientific articles and Wikipedia. The contributors may register any kind of reference. All references are sorted alphabetically in a register making it easy to find other person from the same source. It is important that the URL, links and IDs are stable, otherwise, it is of limited use and should not be included.

¹³ Busetnadssoge, Arnfinn Kielland, tilsett.hivolda.no/ak/BSS/Busetnadssoge.html

¹⁴ www.lokalhistoriewiki.no

¹⁵ www.eidsvollsmenn.no/

Each entry of a person in a source from the first three categories above, denoted by person source entry, PSE, is a separate unit in HPR and is available in the search in HPR. A PSE is a piece of the puzzle that we know should be connected somewhere. Each PSE is only registered once in the database in order to avoid any doublets. Whenever possible, each PSE has a link to the transcribed source in the Digital Archive or in a similar database and ideally also a link to an image of the source. This makes it easy to control the source. If a name differs between the sources for the same person, it is easy to check if this is an error in the source, in the transcription or the name has changed during the person's life time. Since we allow any kind of references, a reference is not sufficient to be registered as a PSE. This would lead to doublets of PSE and poor data quality. We try to include as many high quality sources as possible using the first three methods listed above. We have started an ambitious plan to transcribe all the church books and many other sources, and include these in the Digital Archive the coming years.

Linking

The main challenge in HPR is to link all the PSE regarding the same person. Most persons in HPR have many PSE that are independent of each other in the database until they are linked together. Family relations from the source are maintained. Hence, if we are able to link the baptism of the same person as a child and as a parent, we are able to connect three generations. Family relationship in censuses is slightly more complicated. We use a program from The Minnesota Population Center in Minneapolis that identifies the family relations from the role in the census.¹⁶ However, the errors in the interpretation are slightly more frequent than we expected and we will try to improve this. Since we have internet links to the transcribed source, it is easy to check it manually.

We try to link as many PSE as possible automatically. The Norwegian Historical Data Centre at University of Tromsø has developed automatically linking programs for many years. By August 2016, they have found 1.6 million links between and within censuses and church books. The automatic linking program compares PSEs based on names, names of partners, age/year of birth, household address and municipality. Names are compared using Jaro-Winkler making it possible to identify the same name with slightly different spelling. This automatic linking is performed in a separate Oracle database and the links are imported into the HPR database.

HPR generates an internet page for each PSE and when PSEs are linked, the information is merged and shown on one page. Linked PSE denotes both a single PSE and a group of linked PSEs. The internet page lists the most important information from the sources for the linked PSE and has links to the transcribed sources in the Digital Archive/HPR PSE database. The page shows the names, date and place of birth and death, parents and siblings, partners and children, list of linked PSEs, references and a text as depicted in Figure 1. The text may be a biography and/or it may explain the linking. The list of PSEs will ideally include the main information from baptism, confirmation, marriage, baptism of own children and as sponsor

¹⁶ Sobek, Matthew and Kennedy, Sheela. *Identifying Family Interrelationships in the World's Largest Census Microdata Collection*. Int. Statistical inst. Proc 58th World Statistical Congress, Session CPS070, Dublin, 2011.

and death in addition to censuses. It may also include other events as tax records, prisons, emigration, immigration, events from newspapers, etc.

Busetnadssoge¹⁷ is a program that is used for making a population registers for municipalities. We will also import links from this and similar databases. The critical part is that the links are identified by the correct IDs for the PSE in addition to the general requirement that the links have high quality. Such databases may include persons without any PSE since they are identified by sources that are not transcribed or not given PSE from the included sources. These persons may be included in the separate person input table as a part of the HPR database described in the Section Input data.

Christian Michelsen

Mann
Født: 15.03.1857, Bergen
Død: 29.06.1925, Bergen

[Gå til søket](#)

Partner og barn

Partner: [Johanne Benedikte Boalth Wallendahl](#)
Barn: [Benny Wallendahl](#)

NR	Dato	Kilde	PFID	Rolle	Navn	Fødselsdato	Fødested	Bosted	Fam. St.	Sivst	Yrke	Far/brudgom	Mor/brud	Partner	Barn/andre
1	31.12.1865	FOLK, Bergen, 1865-1865	pf01038249015898		Peder Christian Michelsen	1857	Bergen	Bergen: 14-29	søn	ug					
2	31.12.1875	FOLK, Bergen, 1875-1875	pf01052295014657		Christian Michelsen	1857	Bergen	Bergen: 14-29	Søn	ug	Student				
3	04.11.1881	MINI, Bergen, 1867-1882	pv00000001121068	brudgom	Peter Christian Hersleb K Michelsen	15.03.1857					Uk overrettsagfører			Johanne Benedikte Boalth Wallendahl	
4	31.12.1885	FOLK, Bergen, 1885-1885	pf01053295001211		Christian Michelsen	1857	Bergen	Bergen: Alleegade 59			Overrettsagfører				
5	22.06.1887	MINI, Bergen, 1885-1900	pd000000007656353	far	Peder Christian Hersleb K Michelsen	1857					Overrettsagfører			Johanne Benedicte Boalth Wallendahl	Benny Wallendahl
6	01.01.1891	FOLK, Bergen, 1891-1891	pf01052994015943		Christian Michelsen	1857	Bergen	Bergen: Allegaden 59	hp	g	Skipsreder				
7	01.12.1910	FOLK, Fana, 1910-1910	pf01036693004940		Christian Michelsen	15.03.1857	Bergen	Fane: Gammelhaug	hf	e	skipsreder				

Merknader

Christian Michelsen var skipsreder og norsk statsminister i perioden 1905-1907.

Referanser

- [Wikipedia, Michelsen, Christian](#)

Figure 1. Part of the person page for Christian Michelsen that was prime minister in Norway. The sources are from top census 1865, marriage 1881, census 1885, father in baptism 1887 and censuses 1891 and 1910. The blue text are links to external sources and person pages for family members. The occupation is first as a student, then lawyer and finally as a shipowner.

In HPR, it is possible with manual linking by contributors as crowdsourcing on the internet. Crowdsourcing is particularly important for migration since it often requires manual work for each link. It is necessary to be registered in order to be a contributor. We have started with a small group of contributors and will gradually increase the number of contributors when we

¹⁷ Busetnadssoge, Arnfinn Kielland, tilsett.hivolda.no/ak/BSS/Busetnadssoge.html

are sure that we are able maintain a high quality of the contributions. All contributions are registered by the contributor and the time stamp of the contribution. It is possible to identify all contributions by each contributor. This makes it possible to encourage persons with many contributions, but also to exclude contributors that violate or have contributions of low quality. We also need to get experience with a large number of users and contributors in our HPR data program. The open part of HPR is available on the internet, but we will not massively recruit new contributors before we have more experience. We expect a really large number of users and potential contributors when we gain some experience. Genealogical research is a major activity on the internet.

HPR uses the same search program and data as is used in the Digital Archive. But in HPR this includes also hits in HPR database. The result of a search is a list of data with the vital information from the linked PSEs that match the search. See Figure 2. From each search result, it is possible to put a linked PSE into a basket. This is similar to a basket/cart of an internet shop. In an internet shop, the basket/cart contains the items that you plan to buy. In HPR, the basket contains the linked PSEs that the contributor wants to link, i.e. state that this is information about the same person. From the person page, it is possible to put the person into the basket or perform a search for similar persons.

Søk

PFID	Navn	Kjønn	Rolle	Fødselsår	Fødested	Bosted	Stilling og stand	Hendelsesdato	Kilde	Kildekategori
pf01053257046634	Roald Amundsen	m		1872	Borge	Kristiania: Uranienborgveien 9	Skolediscipel	1885-12-31	FOLK, Oslo, 1885-1885	ft
pf01036371002571	Roald Amundsen	m		1872	Borge Smålenene	Nesodden: Uranienborg	hf ug kaptein, polarforsker	1910-12-01	FOLK, 0216 0217, 1910-1910	ft
pf01037045153788	Roald Amundsen	m		1872	Borge i Smålenene	Kristiania kjøpstad: Professor Dahls Gade 45	ug Sømand	1900-12-03	FOLK, Oslo, 1900-1900	ft
pf01052055035483	Roald Amundsen	m		1872	Borge Sogn	Kristiania: Uranienborgveien 3		1875-12-31	FOLK, Oslo, 1875-1875	ft
pe00000000554744	Roald Amundsen	m		1872		Oppegård	u Forsker	1922-03-17	EMIP, Oslo, 1867-1930	em
pe00000000561184	Roald Engebret Amundsen	m		1872		Oppegård	G Skibsf.	1924-09-27	EMIP, Oslo, 1867-1930	em

Figure 2. Illustration of the result of a search for Roald Amundsen in HPR. The hits are from top: municipal census from Kristiania from 1885, national census in 1910, national census in 1900, municipal census from Kristiania in 1875 and then two registrations from the emigration register from 1922 and 1924 respectively. Occupation is as a student, sailor, researcher and captain.

In addition to linking the PSEs in the basket, we may instead register the link as a family relation, a linkage candidate or a negative linkage. Registration of a family relation may be based on a manual interpretation from a census or knowledge that we are not able to document from a source which is included in HPR so far. Linkage candidates are PSE that may be the same person, where we are not sure. We may establish negative linkage between PSEs from different persons that have similar attributes. Negative linkage gives information to automatic search routines and other contributors, that these PSEs should not be linked. All information regarding a linked PSE is showed at the person page.

All links are given a score between -1 and 10. Score 10 means that all the compared attributes are the same and score 0 means that none of the attributes is the same. So far, we are not able to utilize the frequency of the different names. Score -1 means that there is a conflict, e.g., for a link between two different PSE from the same census. We know that there may be errors in the sources and that the same event may be registered twice. Hence, links with score -1 may be correct.

We apply star linking of the linked PSEs where a lead PSE is linked to each of the other PSEs. When we merge two linked PSEs, all links are reestablished as links between the lead PSE in the new linked PSE and each of the other PSEs. The scores for the direct link between a PSE and the lead PSE is set equal to the lowest score from the sequence of links connecting the PSEs.

IDs

All PSEs in HPR are given a unique ID. The Digital Archive assigns a unique, non-informative 16 digit ID to all PSEs in their archive. HPR assigns a unique, informative 20 digits ID to PSEs from newspapers and similarly for other PSEs in the HPR PSE database. The linked PSE inherits its unique ID from the lead PSE. A contributor merging two linked PSEs, decides which of the lead PSEs in the two linked PSEs that becomes the lead PSE in the new linked PSE. When we search for persons, each hit on a PSE is redirected to the lead PSE in the group that includes all the information from all the PSEs in the group.

References

Use of references is a method to utilize any other kind of information that is available on the internet, in books or archives or any other places but not included in the Digital Archive and the HPR database. This opens for very important and relevant information from a large number of sources. It may also give information of poor quality since we don't want to make strict regulations and are open for a diversity of sources. We don't include private gedcom files with family records into HPR. This would give us a large number of links of high quality, but also many doublet and data of poor quality. But we encourage references to family records on the internet or in books. Here, we may find links and family relations that we have not included in HPR so far. Since we want to document everything in HPR based on high quality sources, we need references to all other kind of information where the quality and completeness of the data is more uncertain. These other sources may have references to primary sources that we don't have access to in HPR and correct links that we are not able to

document in HPR. We expect the users of HPR are able to handle this multitude of sources of varying quality.

All references are listed alphabetically in the register. This makes it easy to find references to the same source. This may be different persons in the same family record, passengers on the same ship or known persons with a biography in Wikipedia. Then we are also able to identify if two different linked PSE in HPR both refer to the same person in, e.g., an American census. Hence, we are able to avoid doublets references to the same source almost in the same manner we avoid doublets of the PSE in HPR. When many persons have contributed to the register, the register will give a good overview over the multitude of possible sources. We will make HPR as transparent as possible by encouraging the use of other sources with proper use of references and with links whenever this is possible. We don't want to replace the other sources. As an illustration we give the following list of possible references:

Geni.com¹⁸, Astri Brun
*Eidsvollsmennenes etterkommere*¹⁹ Jacob Aall eftk.: Nils Hofman Aall
Heimen, 2011, b48. Eide, Thorvaldsen: Andreas A. Svalestuen
Lokalhistoriewiki: Nansen, Fridtjof
Vindern historielag, Medl. blad., 2015/1, Sars: Fredrik Barth
Wikipedia: Nansen, Fridtjof

A challenge here is to find a unique way to write the different references. If the same reference is written differently, we may get duplicated references in HPR. Hence, the list above is important also as an example on how to write references.

Migration

50 million European emigrated in the period 1820-1930 and Norway had one of the highest emigration ratios in Europe.²⁰ The large emigration waves began in the 1860s. Overcrowding and a shortage of farmland in Norway combined with poverty, oppression and class division caused people to break up. Rumours of cheap farmland in America were tempting. Where the emigrants came from varied with time. The large quantities came first from the countryside, but eventually came also many emigrants from towns. In the period 1836-1865, the counties Telemark, Sogn and Fjordane and Oppland had the highest emigration rates. In the period 1866-1915 the counties Oppland, Vest-Agder and Rogaland had the highest emigration rates. About half the emigrants used Oslo as the port followed by the other major cities Stavanger, Bergen and Trondheim. Some of the number from the sources: In Oslo, about 470 000 emigrants were registered. About 20 000 persons that migrated back between 1910 and 1920 have been registered in the 1920 census. There are 114 000 Norwegian nationals in the American 1870 census. In the HPR, we will document both the international and the national migration.

¹⁸ Geni.com, a genealogical database with 89 mill. profiles world wide

¹⁹ www.eidsvollsmenn.no/

²⁰ The National Archive, Emigrant protocols,

<http://www.arkivverket.no/arkivverket/Arkivverket/Statsarkivet-i-Oslo/Norges-dokumentarv-Statsarkivet-i-Oslo/Emigrantprotokoller-fra-Oslo-1867-1966>

There are many sources for migration from Norway.²¹ The police in the port cities made registers on all emigrants from 1865 which is included in the Digital Archive. In the period after 1813, the priest in each parish made a register over migration in/out from the parish often without knowing the final destination. The Norwegian censuses from 1910 and 1920 have information about migration. The American censuses have information on place of birth for persons and in the period 1900-1930 also for the parents. These sources are part of or planned to be included into the Digital Archive. There are also several other sources like passenger lists from ships. None of these sources are complete, but together they make it possible to link a large number of persons giving a better understanding of the major migration from Norway in the period 1825-1940. There will be a close cooperation between the development of HPR and the international cooperation in the North Atlantic population project, NAAP.²² Hopefully, we will be able to include the links made in the NAPP-project between Norwegian and American censuses.

Other sources may be used as references. This may, e.g., be links to regional population registers in Sweden²³ or Denmark.²⁴ Typical references on migration when persons also are linked to Norwegian sources:

Danmarks Adels Årbok, Bugge: Christian Bugge
Ellis Island, Kristianiafjord 1913: Henrik Hansen (Immigration to US)
Odensedatabasen: Peder Hansen ID 190001, d. 1743
Rotemannarkivet: Ole Andreas Edvard Olsen b. 1856-09-26
WeRelate.com²⁵: Karelius Olsen (1)

There are many local population registers covering a municipality. The local database for Rendalen, a remote municipality with limited mobility, has a linkage rate above 90%. At places with more migration like the mining community Røros and larger cities like Oslo, it is not possible to obtain the same linkage ratio. Persons that live at the same farm in their entire life and families that remain in the same municipality for generations are easier to follow than persons and families that migrate. Continuous or stepwise migration is more difficult to follow than a single migration. It is easier to follow persons that marry within the local community than persons that find their partner outside the local area. Persons that don't have a permanent address are, of course, even more difficult to follow. Hence, all population registers have a bias due to poorer coverage of migrants.

Migration is a major motivation for a national population register. We want to make statistics for the migration between the different municipalities in Norway and other countries for each decade. We may also analyse the sex, age, the order of sibling and other variables for the migrants. This statistics will be based on counts for persons where we have identified the migration and will gradually improve as the register becomes more complete. The migration will be identified both from registration of the migration, e.g. emigrant records, and for

²¹ Eide, Ole Jone and Thorvaldsen, Gunnar, *Historisk befolkningsregister: Oversikt over emigrasjonen*. Heimen (ISSN 0017-9841). 48, 2011.

²² Thorvaldsen, Gunnar. *Using NAPP Census Data to Construct the Historical Population Register for Norway*, *Historical Methods*. 44 (1) 2011, s. 37-47.

²³ e.g. Rotemannsarkivet

²⁴ e.g. Odensedatabasen

²⁵ WeRelate.com, a genealogical database with 2,6 mill. persons, Minnesota, US

persons with events in the two different municipalities. In the latter case, we will assume that the migration is performed at some time between the two events. In both cases we must handle missing information about e.g. the origin of the migration and the order of siblings and that the same migration may be registered in several different sources. It is more difficult to make statistics for stepwise migration and back migration. We need more data and experience before we are able to specify this statistics. As a first step, HPR documents many examples on migration.

A national register makes it easier to follow migration than local or regional registers. Place of birth is specified in most sources. Migration inside Norway is documented by linking PSE for baptism, death and possible censuses. Migration outside Norway is documented by linking PSEs for emigration, immigration or from international sources with other PSEs. In principle, migration in and out of Norway should also be registered at least once and often several times. This gives us PSEs that document migration. When we have made all the obvious links in HPR, it will be easier also to find the more complicated ones. Linking candidates may be used when we have found a likely link but when we cannot be sure. Strays are persons found in one source that have migrated and that will not be easy to find from the place where the person migrated from. Currently, there are list of strays that may be used in linking.²⁶ With a national register, it will be possible to link the stray to other sources directly instead of making separate lists.

The strategy is clear. We will try to include as many primary sources into the Digital Archive or directly as part of HPR as possible. This includes sources with passenger lists, emigrant records, Norwegians in other national censuses, etc. It may be a transcript of the original source or documented in a newspaper. Sources that are included in HPR will be available through a search in HPR and are piece of the large HPR puzzle that we know should be connected somewhere.

Research and the use of HPR

It is important that HPR is able to attract the scientists and other experts (employees at museums, genealogists, etc.) and make them contribute to HPR. Then we need to make HPR valuable for them. The main purpose of HPR is research. This is where we have obtained the major funding. Hence, the focus will be to give information in a form that is valuable for research. We will provide statistics and graphics that will cover the most typical needs of the users. But this output is more likely to satisfy an average user of HPR than an expert. The expert will probably prefer to export data from HPR into their own private database. Export of data is a central feature of HPR. For HPR, it is important that this use also leads to improvement of HPR itself. We will encourage expert users of HPR to improve HPR before data is exported. We are also able to import links from other programs into HPR where we know that the links are of high quality. The imported links need to refer to unique IDs provided by the Digital Archive for each PSE. HPR may also be used as documentation of research result since it may be document historical persons and events and have permanent IDs for the entire historical population.

²⁶ DIS-Norge, strays, <http://www.disnorge.no/slektsforum/viewforum.php?f=2381>

For HPR, it is also important to be attractive to the large population of hobby genealogists. Many of these persons have a thorough knowledge of the sources. This group is likely to provide the major part of links in HPR. We would like to attract persons with a variety of interests: local history, following a family or farm, particular groups of persons or associations, etc. We will establish close relationship with the relevant associations in order to make HPR a valuable tool for these persons and associations and to promote contribution of high quality to HPR.

The database structure in HPR.

HPR is using a MySQL relation database where the most important tables are:

Person units (PU) with ID, name, birth and death date and place, biography, link references.

Person links with the ID to PSE and PU, timestamp, score and contributor

Family relations with IDs to PSE, family relation, timestamp, contributor.

References text, URL, timestamp, contributor.

We only store family relations that are not specified in the source. Both person links and family relations may be negative, specifying that these persons are not the same/or not a given family relationship. There will be some graphics and we will expand the statistics in HPR, but this is not determined so far.

Acknowledgment

The authors thank the other partners in the project in particular Gunnar Thorvaldsen, the Arctic University of Norway, Lars Nygaard, the National Archive, Kåre Bævre, the Public Health Institute and Jørgen H. Modalsli, Statistics Norway. The authors also thank the Research Council of Norway for funding the project under contract no 225950 as part of the national infrastructure program.