**Record linkage in the Historical Population Registry for Norway**[1]

*Trygve Andersen and Gunnar Thorvaldsen, Norwegian Historical Data Centre, University of Tromsø*

Construction of the Historical population register (HPR) is now ongoing, nationally for the period 1801 to 1815. Our first goal is to reconstitute the Norwegian people on 17 May 1814 when the constitution was signed, a part of the upcoming centennial celebrations. In addition we aim to construct a nominative census to supplement the statistical census of 30 April 1815. This will be the world's first widely available and nationwide population registry and cover the period 1801-1815 – although record linkage rates will be significantly under 100 %. This project has been named "The Norwegian people in 1814" (DNF1814), and is implemented by representatives of the National Archives, the Norwegian Historical Data Centre at the University of Tromsø, the Norwegian Computing Center, the genealogical society "DIS-Norway" and other volunteers (Bråthen 2011). As supplements to the1801 census the church records 1801 to 1815 have been transcribed and record linkage is progressing. However, DNF1814 represents only the initial part of the HPR project, which recently became realistic through substantial funding from the Research Council of Norway. The HPR aims to cover the entire Norwegian population until ca 1920, since the Nordic countries have relatively complete vital registers filling the gaps between the censuses (Thorvaldsen 2008 and 2011). We also plan a restricted access database for the period until the Central Population Register (CPR), which was created in 1964. These two restricted access databases from 1920 and 1964 respectively will not be dealt with in this paper, which particularly focuses on record linkage between the 1801 census and the church registers 1801 to 1815 and other parts of the open period.[2]

The HPR will in principle link all openly accessible information about historical persons and their locations in Norway. It will be built mainly on church records and censuses, but may eventually include information such as emigrant lists, newspaper notices, prison records and tombstones. It will link information about the same persons appearing in several sources and their settlements. This is an enormous task that would never be finished if all potential source material should be taken into consideration. However, our ongoing project's goal is to transcribe and link the censuses and church records during the next ten years. The period from 1735 to 1964 includes 9.7 million people and more than 37 million source events in censuses, church records and emigration protocols. The Research Council funding should last from 2014 until 2018 with a focus on constructing the necessary software and other infrastructure, start filling in data more efficiently and complement with the rest of the main source material during the next five year period.

**New opportunities with the HPR**

The DNF1814 database and later extensions of the HPR will be used in a variety of local, regional and national studies and provide a basis for international comparisons. The HPR includes international collaboration through the North Atlantic Population Project (NAPP) and the European Historical Population Samples network (EHPS-net). It will give us a new historical and social science understanding of the relevant periods. With longitudinal microdata we can study e.g how family structure and social and geographical mobility changed longitudinally as opposed to the snapshots given in the censuses. In a medical perspective, the HPR will be an important source for studies of the Norwegian populations gene pool and for instance genetic diseases related to consanguinity. The bibliographies maintained by the Demographic Database in Umeå and the Minnesota Population Center in Minneapolis contain a host of further research topics that have not yet been studied in Norway.[3]

---

[1] We are grateful for contributions from other team member, especially Lars Holden and Torkel Bråthen.

[2] Cf further information at the projects web pages: http://www.rhd.uit.no/nhdc/hpr.html and http://www.arkivverket.no/arkivverket/Digitalarkivet/DNF1814

[3] Cf http://www.nappdata.org/napp/, http://www.ehps-net.eu/, http://www.ddb.umu.se

Within local history and genealogy it will be easier to place people's family and community history into a broader context. In practical terms, it will be easier to identify the sources and more efficient to link to other people's work on their ancestries. It will always be possible to find more information, more sources and comparable life histories. It will be less necessary to retrieve the same individuals and duplicate the same links, and we can instead complement the work of others. The aim is a database where we can follow individuals, families, farms, homes and other locations over time. It is not trivial to establish such a population registry, and this paper describes some of the challenges we face and how they are solved. Record linkage will be done both with automated algorithms and interactively via the Internet. Thus, the HPR represents new technology for collaborating to link personal data.

**Main principles**

The HPR will mainly rely on the input of data from censuses and church records, ensuring that we include all person records in these sources only once. The main principles behind the automatic and manual linking are the following:

1 The HPR will build on many available sources of good quality. There should be two-way pointers, that is, from the database to the sources and vice versa.

2 We want the greatest possible openness and transparency. This makes it possible to see who made the links and the criteria applied.

3 We encourage the greatest possible development of high quality links. All users are given opportunity to comment on database quality.

4 All links will be marked with quality and linkage rule flags, and the representativety of linked samples can be compared statistically with the population in the decennial censuses since 1815.

These principles ensure uniqueness and source references. It will ensure increasing data quality over time. The HPR functions also as an index to the source entries instead of replacing them. With regard to citations and transparency, the HPR will be different from the historical population registry in Iceland, deCode Genetics, which is closed so that people only have access to their own ancestry and is somewhat lacking with respect to source references and criteria for record linkage. The historical, longitudinal databases in Sweden and the Netherlands have a better basis for linking in their pre-linked, original sources, but so far cover only parts of these countries.[4]

**Basic source material**

In Norway church records started gradually from 1623 with entries about baptisms, marriages and burials; - their keeping has been compulsory since 1680. We expect, however, an undercount of 10-20 % in the 1700s, particularly of children's burials. This is one of the reasons why the HPR will start in 1801 with the nominative census of that year as their starting point, but it also because the earliest sources contain poorer information; ages may be missing, names of married women are dropped in several events and residence data may not be available. From the period 1800 to 1960 some ten million out of 30 million records in censuses and church records have been transcribed.

The National Archives now establish unambiguous identities (PIDs) for each person in the transcribed sources as part of its Digital Archive.[5] These IDs are used as references in the HPR. This makes it easier to determine whether there are additional or conflicting information available from the same source. Combination of many sources will improve the HPR both by providing more information about people and places, and make linking more reliable. Complementary sources can be anything from probate protocols to newspaper reports.

[4] http://www.decode.com/; Cf http://www.decode.com, http://en.wikipedia.org/wiki/DeCODE_Genetics, http://www.iisg.nl/hsn/

[5] Cf http://arkivverket.no/eng/content/view/full/629

## Processing of data

The censuses contain information about the persons' position in the family and household, which provides the basis for connecting individuals. The HPR establishes the family structure automatically by creating location variables based on information about family position, order in the residence, gender, age and name. This system employs algorithms developed by our partners at the Minnesota Population Center as part of the North Atlantic Population Project (NAPP).[6]

The source entries have been transcribed verbatim, faithfully reproducing the names, year of birth etc. The same person may thus have variant spellings of names, different age data etc in different sources. For the overall presentation of each person's data we choose the core version of variables in which we have most confidence. The core version is selected based on the following priority rules: 1) manual selection, 2) Church records where the older takes precedence over newer entries, 3) censuses where later take precedence over older and 4) other sources where later take precedence older. Among church records the oldest will be the baptismal entry, whereas among the censuses the quality improved over time.

**Record Linkage and Coupling**

*Linking* can be defined as identifying and determining that the same person is referred to in at least two different sources. *Coupling* is tying together various family members and is based on sources that mention several people in the same event, for instance baptism or census taking. Linkage and coupling are interrelated processes, since family information can be used to link persons and information from several sources may be necessary to decide what persons are related. Several genealogical databases now use MediaWiki software to link and couple records via the Internet. For instance WeRelate.com is the largest genealogy wiki with over 2.5 million people registered in the database. These are largely based on reading the finished family trees via gedcom files from genealogists. Since they are not based directly on the source material, they thus create problems with duplicates and database quality. We have attempted both to use the WeRelate platform and to construct our own wiki version, but have decided that building a relational database gives a more flexible and efficient solution.

The size of the HPR project makes this necessary. In the HPR we read and store all the events from the transcribed sources. This places great demands on the system as transparent, simple and designed to strengthen database quality over time rather than degenerating. A key element in manual record linkage is how to motivate many users to provide high quality links and how these contributions are monitored and the quality is assessed. It must be easy to find all the people featured in a specific source, all families in a municipality, all priests in Norway or other groups of individuals, families and places.

The national objective makes it possible to capture migration between different areas and emigration abroad. People who move between areas can usually be retrieved in both areas and we can establish special lists for migrated people that are not identified. Similarly, it will be possible to rediscover the roots of most people of Norwegian descent living abroad in the HPR since the emigration lists from ca 1870 with nearly one million records have been transcribed. From a research point of view it will be particularly interesting to follow the people who moved and that we have missed in previously reconstituted local databases. Many users will surely enjoy the challenge of linking migrants. Experience from other projects, such as the Local History Wiki,[7] shows that different contributors will share knowledge from their particular expertise. Topics may include own family and residence, geographic areas, specific occupations, ethnic groups, persons mentioned in the literature. It is possible to describe the work within a particular theme in terms of project pages, inspiring users with different approaches and techniques to work on the same data.

[6] These pointer variables and "famsize" – number of members in own family in the household – can be analyzed with other variables constructed by Minnesota Population Center, viz http://www.nappdata.org

[7] Run by the Local History Institute - http://lokalhistoriewiki.no/index.php/lokalhistoriewiki.no:Hovedside

All links will get a quality flag on a scale from 0 to10, viz the following guidelines:
10: completely secure link, criterium 8 pluss identifying the person as part of a couple or family.
8: the same birth date[8] and same or similar names as well as geographical origin on the farm or street level.
6: The same or similar name of both spouses and geographic affiliation.
4: same name, age and related to the same place.
1: a probable link, the records should likely be linked.
0: established as linkage candidate with reasons specified for the uncertainly.

**Automatic or manual linking?**
Only in the DNF1814 part of the HPR project (1801-15) information on ca one and a quarter million inhabitants are to be linked: 888,000 in the 1801 census, the 350,000 births plus an unknown but small number of immigrants to Norway during the first 15 years of the 19th century. This requires significant efforts and as much as possible should be automated. It is well worth noting the opinions of the pioneer Danish demographer Hans Christian Johansen - the Danish sources were equal to the Norwegian during this period before the union was split in 1814.

With reference to Figure 1, there are two competing goals. First, to link as many individual records as possible, and second, to avoid linking entries that do not belong together. The ideal is represented by the origo - bottom left, where all linked entries really belong together and no true links were omitted. This is hardly ever possible to achieve with historical records because of unprecise and missing information. The problem is that when we impose stricter rules in order to reduce the risk of linking records about different individuals (horizontal axis), we also rule out several links that really should have been included (vertical axis). Conversely, if we introduce more liberal rules in order to ensure that all potential links are realized (towards bottom of the vertical axis), we may include more links that really should be discarded (towards the right end of the horizontal axis).

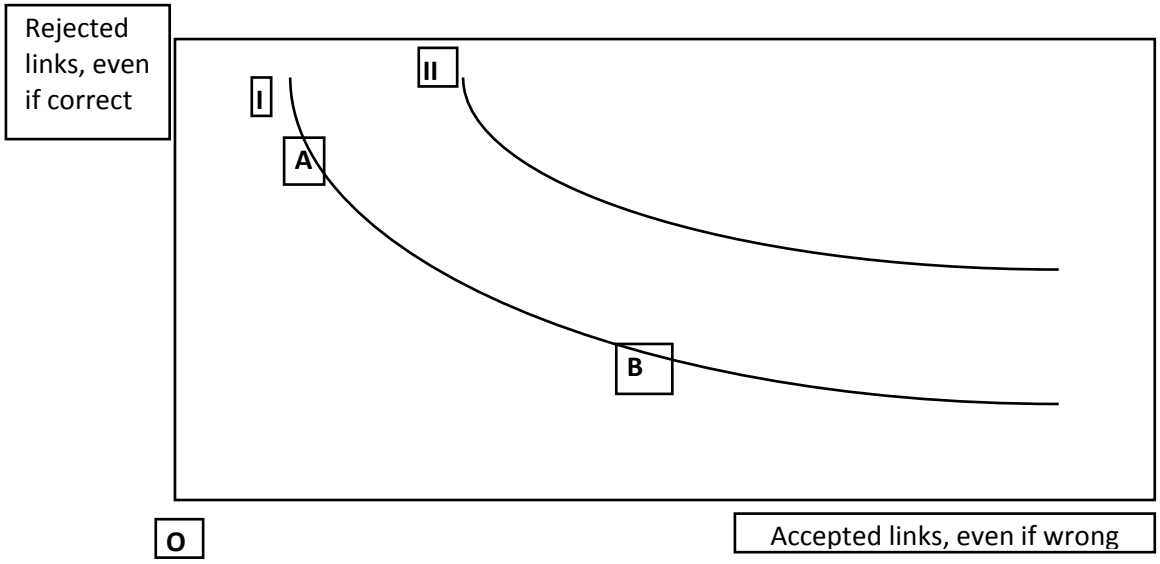


Figure 1 The ratio of accepted links with errors ("false positives") and correct links that were rejected ("correct negatives) in sources from the early 1800s (curve II) and late 1800s (curve I). From Johansen (2002).

Hans Christian Johansen assessed that while getting close to origo is the ideal, acceptable linkage results lie between points A and B on the curve in the diagram, with few accepted erroneous links (A)

[8] Birth dates are found in many church records from 1870, for some persons in the censuses from 1891 and for all from 1910.

and few excluded appropriate links (B). Linkage results along curve II he thought unacceptable because it introduced many false links in the database and many correct links were omitted. His main point in our context was that only from the second half of the 1800s, we have source material with a precision that allows automatic linking along curve I, while sources from the earlier decades are so imprecise that they must be linked manually, otherwise we end up on curve II. Fortunately, we have over the past decade developed techniques and methods that may allow us to conclude differently, at least for parts of the material.

Firstly, we can now dynamically bring together variant spellings of the same name, both using standardized lists of names and algorithms for the comparison of strings. Secondly, we have developed software which not only links individuals, but takes into account couples or entire families when linking. Both techniques are particularly valuable when dealing with a period when information other than names and ages was imprecise or missing from the sources. By combining dynamically name and family information we can shift a significant part of the population to be identified in several sources from curve II to curve I in Figure 1. Analysis of the 1801 census shows that about 80 % of women and 85 % of men were living with at least one other family member. Even if only a portion of the relations were stable over time, it still shows there is much potential in linking with grouped criteria.

In the next round, the problem of balancing the proportion of correct and incorrect links can be reduced further in the HPR by adding manual record linkage. The same quality and methods flag requirements apply to all linking, providing a quality indicator and justification for each link. Thus, researchers can make independent assessments and choose what links they will trust in their analyses. In sum, automatic record linkage has the great asset that documentation is inherent in the algorithms and makes the links easier to trust, especially when using material linked by others. However, there are decisions based on background knowledge and details in the sources which are not easy to automate: For instance: Nick names used in certain families persons known to have changed their names, knowledge about ancestors who cohabitated, names of neighbouring farms that were used interchangeably for cottars' places etc. Lists of property sales and probate registers have been digitized only to a small degree, but have been used by many genealogists to build their ancestral pedigrees.

**Automatic linking: The Parish of Lenvik**

The Norwegian Historical Data Centre has developed software for the automatic linking of married couples and other family members, also utilizing special algorithms for comparing names.[9] In connection with the DNF1814 project we now run these programs to link the 1801 census and the church records for North Norway which were transcribed for the relevant period. Here we use one parish to exemplify. The starting point were the men who were fathers in one or more of the 696 baptisms in the period 1799 to 1815, from 16 (in 1811) to 63 (1806) annually. The histogram in Figure 2 shows how many fathers' records it was possible to link to the different types of other records each year. Automatic linking provided the best results for the baptismal records where both father and mother are brought together year after year. Around two thirds of the parents were linked to the wedding entry, a stable number over time. Both spouses' names are on both lists and the parish priests were consistent with spelling etc. The biggest problem was linking to the 1801 census, for there many couples are found in various places of residence before they married and had children together. Thus, for the 1801 census and burials we typically get good linkage results early in the period and weaker towards the end because a greater proportion of couples had married. The funerals that were linked to the father at baptism are mostly children, which must be understood against the background of the high fertility and child mortality in this period.

[9] Developed in PL/SQL, interface to our Oracle databases. Names are compared with the built-in Jaro-Winkler string comparison algorithm.

If the residence is entered in the parish register, and is the same as in 1801, it is also possible to link individuals with no age data provided in the parish register. Out of 834 funerals, 349 were not linked since especially for older men, it was common to be entered in the parish register without reference to relatives. Access limited to first and last names are often inadequate linkage criteria. Also later in the century many were left unlinked: a third of the people who we try to link between censuses in 1865 and 1875 for portions of Troms with more traditional techniques were not identified in the source (Thorvaldsen 1995). We see, in other words that there is considerable scope for manual linking employing the detailed knowledge of families and communities which exist in abundance in Norway. With the HPR's Internet system, we expect to activate this knowledge and resolve many of the problems individual investigators struggle with in vain.

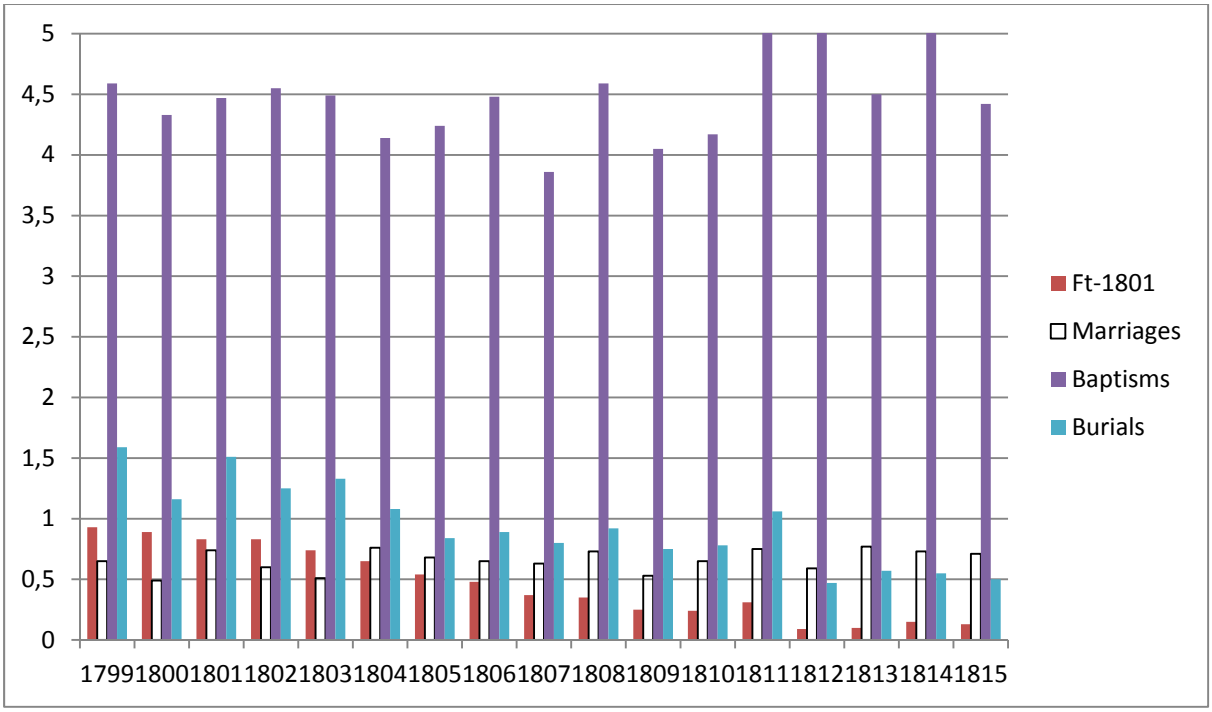


Figure 2: Average number of links from the father at baptism with the 1801 census, marriages other baptism records, and burial list from the period 1799-1815 for Lenvik parish.

**Database Structure**

The HPR is in the process of receiving substantial infrastructure funding. We can therefore afford to leave the ready-made modules of the MediaWiki platform behind and build a relational database with similar but more flexible functionality. The wiki database was programmed in the languages PHP and C, and the upside is a wiki's well known ability to save all changes signed and time-stamped. Thus, it is easy to reverse any or all changes made by any user. There are discussion pages to debate the contents, for example whether a link is correct or not. The downside is the wiki's demand for resources both in terms of disk space and processing power. Even with the partial number of source entries we are now dealing with, users easily get frustrated by the waiting time for each and every transaction detail, not to mention the need to steadily update our hardware. With a relational database tailored to our purposes, we get similar functions, more flexibility and more acceptable updating times.

The main feature of the system is the establishment of a separate page for each person and each family. A *person* page is of the type "Ole Hansen (21)", which says that there are currently 21 occurrences of entries for persons named Ole Hansen registered in the HPR. Figure 3 shows an example of an individual's page. Here you can collect all source information about the person, links to family members and even put a picture or write a biography. A *family* page can be entitled "Ole

Hansen (21) and Maria Thorsdatter (8)." Their residence information is created from censuses and has pointers to other residents. These sequence numbers are only used in the places, families and persons where it is necessary to provide uniqueness. Settlements are usually not established from the church records because these are less precise in their description of objects. The pages constructed from the source entries are merged / linked when we have sufficient certainty that it is the same person, family or residence is mentioned in various sources.

Figure 4 gives an overview of the entities in the HPR. An individual entity is created for each person's occurrence, to be linked to source entries. Family pages are created for each family from a marriage record or one parent with a child. Family pages are merged when the parents' records are linked. A place page for a farm may thus point to different municipalities in the different censuses because of municipality changes through time. In addition, the project page where you can describe the work in a historical society or any other topic. Since the Project will provide links to external sites about the project and links to relevant pages of the HPR, such as specific homes or people. There are also entries about each source used in the HPR.

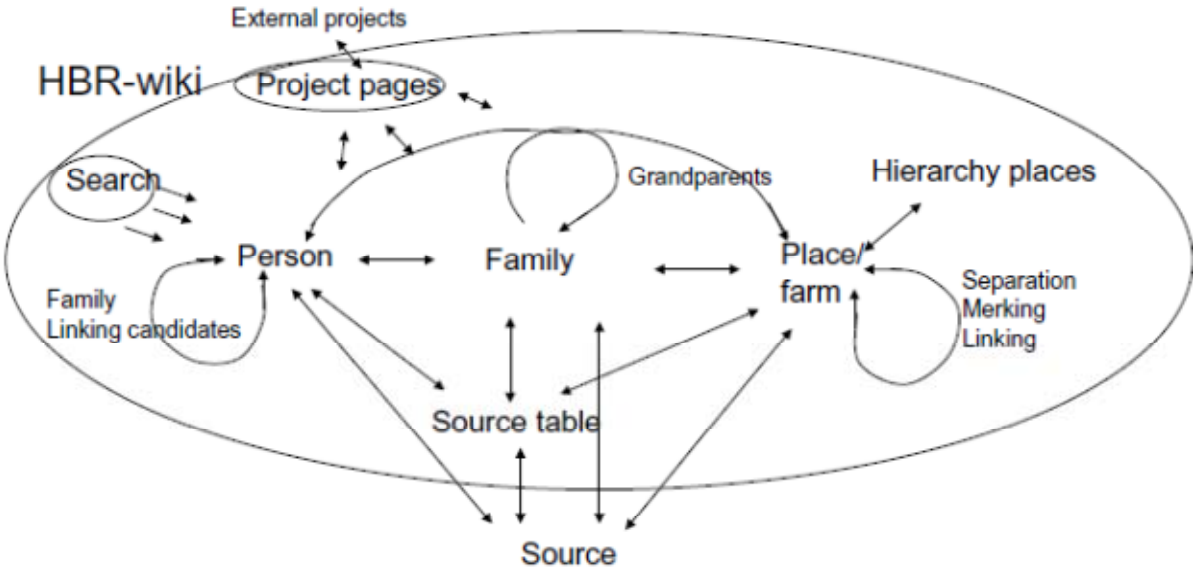


Figure 4: One-and two-way pointers in the HPR represented with single and double arrows.

**IDs for the people**

It is necessary to establish the IDs of all persons in the HPR so that we can have a secure reference to persons within the HPR, and between the HPR and external programs. Today's social security numbers rely on stable and unique identification based on birthdate, gender and additional digits. However, date of birth will often not be known precisely for historical persons and cannot be used as part of the identifier. This is why the National Archives is establishing IDs for all source entries (PKID) and properties (eKIDs) listed in the sources. We consider this as a reliable identification of each individual occurrence, referring each entity to a well-defined source entry. If you later decide that the name or other information about the person is transcribed wrongly, it will not change the PKID.

The HPR uses a rule-based definition of each individual's ID, when the PID is based on the PKIDs. The PID is defined as the PKID with highest priority of links to a particular person. This is a system that can also handle changes in the linking. This is illustrated in figure 5 where the PKIDs A and B were originally linked. However, when B is instead linked to C the latter's PKID takes preference.

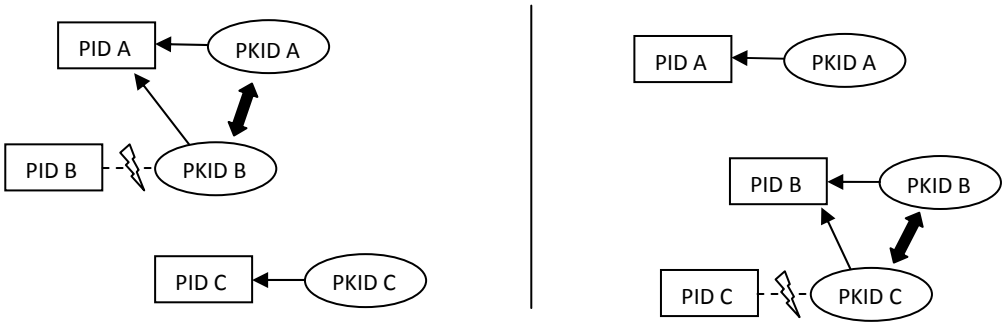


Figure 5: Changing the person's ID (PID) when linking two person records (PKID). In the example on the left the PKID A has priority, and PKID B "inherits" the PID since the records refer to the same person (thick arrow). The example on the right shows that record B is instead linked to record C. The previous link is broken. Record B's PKID becomes the PID because PKID C has lower priority.

In the HPR we use the following priority rules to create the PID from the PKIDs:
1 Census 1910
2 Funerals / lists died before 1910
3 Births after 1910
4 List emigrated before 1910
5 Other censuses where recent censuses have higher priority than older ones.
6 Other entries in the parish registers where older entries (especially baptisms) have higher priority than newer
7 Other sources where more recent entries have higher priority than older ones.

The first five categories are not overlapping, and the first three are presumably of relatively good quality. The 1910 census will later be used to link between the HPR and the closed portion of the population register from ca 1920.

**Input of linked collections**

The results from the automatic linking will be copied as lists of groups of PKIDs (the Ids described above) to be revised interactively. It is necessary that we include references to the sources and identify the individuals consistently and the PKIDs fulfill both requirements. The tables will contain information about the quality of the links on the 0-10 scale (cf above) and linkage criteria. In the HPR we try to get contributions from everyone who has linked major collections of the Norwegian population. We also experiment with semiautomatic linking methods. A standard set of rules must be developed, the links flagged according to rule used and controls run to verify consistency. Thus, as soon as a new source is read into the HPR site, the straightforward links (on the two or three highest levels) have been established automatically or semi-automatically.

One person in the HPR cannot have multiple fathers or mothers. It is therefore not possible to merge personal entries where there is conflict between the parent candidates. Even if a person's ancestry cannot be linked because of a conflict between competing links to the father or mother, there is always hope that the parents' person records can be linked together by finding additional information at a later stage. Thus, children's and parents' potential personal records are marked as candidates for linkage. As already stated, we do not consider reading in GEDCOM files, since this requires that stable and unique source references exist. Otherwise redundancy is easily introduced in the database.[10]

[10] In WeRelate substantial efforts have been made to reduce redundancy. Cf: http://www.werelate.org/wiki/WeRelate:Suggestions/Tweak_to_Duplicates_Report

**Manual linking**

We expect that a significant part of the linking and coupling in the DNF1814 will be done manually with contributions from genealogists and local historians volunteering due to the simple structure of the source material, missing data and lack of consistency in this earliest part of the HPR. The manual record linkage can be informed by checking the HPR against their personal databases or genealogies. Also, they can search for persons' records among the chronological events in the church books, all events on a farm over time or other criteria. If in doubt, they can flag person records as linkage candidates and ask for the opinion of others to confirm or reject the link. Especially in the early church records and for emigrants and immigrants it may be necessary to establish candidate links. Registered coupling, i.e that two people are related to each other such as parents and children, are marked by establishing pointers between person and family records.

**Historic Sites and administrative boundaries**

The evolution of the population is of course closely linked to the development of settlements, communities and regions. It is an ambition in the HPR project to link to this information which is nearly as dynamic as population developments: the creation of smallholdings, the splitting of farms, urban growth etc. A residence may be linked to several municipalities over time if administrative boundaries changed. We can start with the 1838, 1886 and 1950 land tax register and connect census records to these, using farm name, place name, street name, title number, farm number, house number and street number.[11] Results from the project *Historical administrative boundaries* shows that 80 to 90 % of the farms can be linked to a unique title number. The dynamic farm register developed in this project by dr Kåre Bævre can be linked to the HPR. Urban places will require extensive new work due to the transition from the serial property numbers in the city to street names and numbers. Again, we rely on local and regional expertise to solve these puzzles correctly. Historian Arne Solli at the University of Bergen has worked with the person and location information for cities in BerGIS project for the period 1696-1906 and is seeking to extend this GIS to other urban areas, cf http://bergis.uib.no/index.shtml and Solli (2006). Based on the coordinates for farms and urban blocks it should eventually be possible to create illustrative and analytical graphics, such as displaying the migration patterns of a neighbourhood or the spreading of epidemics.

**The Rendalen database (1735-1950)**

We can get an idea of the challenges we face and what kind of results can be expected from longitudinal databases at the local level. In the rather isolated parish of Rendalen on the border with Sweden, church registers, censuses and other sources have been linked for the period from the ministerial records started until the last pre population register census. The lowest linkage rates are in the 18th and early 19th century, before the decennial nominative censuses started in 1866. Of all the people that have been observed in the period 1815-1824, half were identified in both baptism, marriage and burial records, while another 20 % are identified in baptism and burial, but not in marriage. In view of the high mortality in this period we expect that a large number died unmarried. Rendalen was a parish with little migration and the database is the result of thorough manual work on a small geographic area. We cannot expect equal coverage in municipalities with larger migration, but Rendalen can function as a "gold standard" to facilitate linking of other paprishes.

Even after we have linked using the probate records, as has been done in Rendalen, we face challenges in order to reconstitute the population at a given time. These challenges are not diminished by the fact that mortality was high and that part of the population was geographically mobile during the turbulent times in the early 1800s. The high mortality meant that over one hundred marriages where recorded with widows or widowers involved in Rendalen from 1801 to 1815. To determine how these settled together with children from previous marriage over the

[11] Cadaster 1950: http://www.dokpro.uio.no/cgi-bin/stad/matr50, cadaster 1886: http://www.rhd.uit.no/matrikkel/excel.html. The cadaster from 1838 has been scanned and can hopefully be made available in machine-readable and searchable text form after being processed with OCR.

following decades is no simple puzzle. The next nominative census was taken in 1866, but some help is rendered by the silver tax lists from 1816 and the farm tax lists from 1838 – at least for the heads of households.

**Summary**
A national population registry will become a unique historical source for the last two centuries and may be used in many different research projects. The potential inherent in the rich Nordic source material will be released once the nominative records are linked together in order to describe persons, families and places longitudinally. This requires the development of new linkage techniques combining both automatic and manual methods, consisting of a composite of several established techniques combined with new methods that increase the overview. It will require a large concerted effort from a large number of people to achieve these goals, but in return this gives us an infrastructure with great utility for many different research projects. We are cooperating with several international projects in the field of computerizing nominative sources, and look forward to future exchanges of ideas, especially on how to better use combinations of the church registers in population research. From an international perspective, there are no comparable open and national historical population registers built by linking multiple source types.

**Notes**

## Litterature

Bråthen, Torkel Rønold. «Det norske folk i 1814» i *Slekt og Data*, nr 4/2011, s. 44-45.

Bull, Hans Henrik. *Data and Methods*. Universitetet i Oslo, 2006, s. 25-34.

Drake, M. *Population and society in Norway 1735–1865*. Cambridge, Cambridge University Press. 1969.Eikvil, Line, Holden, Lars og Bævre, Kåre. *Automatiske metoder som hjelp tiltranskribering av historiske kilder*. Notat SAMBA/44/10, Norsk Regnesentral. Oktober 2010.

Gjelseth, M. *Relasjonsdatabaser som verktøy i en historisk-demografisk studie*. Hovedoppgave, Historiskinstitutt, Universitetet i Oslo, 2000.

Heimen, spesialnummer om personvern, 2005 med blant annet With, Christian. «Slektsforskning og personvern» i *Slekt og Data*, 1/2005. Tilgjengelig på Datatilsynet sider, http://datatilsynet.no/templates/Page____1356.aspx

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

Johansen, Hans Chr. «Identifying people in the Danish Past» I Sølvi Sogner, Hilde Sandvik, Kari Telste, Gunnar Thorvaldsen (red): *Pathways of the past: essays in honour of Sølvi Sogner on her 70th anniversary 15. March 2002*. Novus, Oslo 2002, s. 103-110.

Arne Solli. *Urban space and household forms*. Artikkel presentert på the Eighth International Conference on Urban History, Urban Europe in Comparative Perspective, Stockholm 30th August – 2nd September 2006.

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

Thorvaldsen, Gunnar. «Fra folketelling og kirkebøker til norsk befolkningsregister» i *Heimen*, bind 45, 2008, s. 341-359.

Thorvaldsen, Gunnar. *Migrasjon i Troms studert i folketellingene 1866-1900*, Universitetet i Tromsø 1995.