

12 Norway

The Norwegian Historical Data Centre

Gunnar Thorvaldsen

Introduction

Researchers and archivists in Norway have created several databases containing transcripts of historical, nominative sources such as censuses and ministerial records, continuing a century-long tradition of record-keeping that tracks virtually every Nordic citizen. This article describes the Norwegian Historical Data Centre, a permanent part of the University of Tromsø, which transfers the information from nominative material—such as censuses and ministerial records—into digitized form.

The Norwegian Historical Data Centre

Since the seventeenth century Norwegian bureaucracies have created vast amounts of records on the individuals and families they administered. The most central are the censuses from the 1660's onward; ministerial records cover some areas even

Dr. Gunnar Thorvaldsen is professor of history and manager of research at the Norwegian Historical Data Centre, at the University of Tromsø, Norway. His main research interests are historical migration, mortality studies, and the computerization of historical microdata. In 1996-97 he spent a sabbatical in the History Department at the University of Minnesota, Minneapolis, and participated in the mortality study group at the Center for Advanced Study in Oslo during 1998-99.

For more information on the Norwegian Historical Data Centre, see the web at:

<http://draug.rhd.isv.uit.no/rhd/indexeng.htm>.

earlier. There also are probate records and judicial material from roughly the same time, while records dealing with special groups like sailors or emigrants are of more recent origin. The quality and geographical coverage of these nominative sources vary over time and from country to country. The union of Sweden and Finland on the one hand, and Denmark and Norway on the other, created different types of source records which were maintained within each of the twin countries even after the unions were dissolved in 1808 and 1814, respectively.

For three decades, historians and other researchers have made the material more available to scholars with the help of modern computer technology. This article outlines the major databases containing historical records in Norway; other chapters of this book outline similar databases in Sweden and Denmark. The focus here is on resources built to serve the general research community. Databases built for specific research projects fall outside the scope of this article, which deals primarily with nominative, individual-level data.

The first major project was begun at the Historical Institute at the University of Bergen which computerized the 1801 census along with parts of other censuses, emigration records and ministerial records in co-operation with the Regional Archive. Today they cooperate closely under the auspices of the National Archivist. The Historical Institute at the University of Oslo dealt with similar material from the capital and its surrounding area within the framework of a project producing more than thirty theses on social science history and demography (Langholm 1976). This Institute no longer enters historical material into the computer, however, in large part due to the establishment of the Norwegian Historical Data Centre (NHD) at the University of Tromsø in 1978. Since 1985, the NHD has been a permanent body within the School of Social Sciences, serving researchers, teachers, students and genealogists nationwide.

The long-term goal of the NHD is a national population registry of the eighteenth and nineteenth centuries, designed primarily for research purposes. To date the NHD has computerized more than 1.5 million records from censuses and an additional half a million records from the Norwegian parish

registers of baptisms, births and burials. This is a long and rich series of records which, in many places, extend back to the 17th century, so the bulk of these ministerial records are still not computerized (consult the church books available from the Digital Archive). Work has begun on the probate registers, constructing inventories that help researchers find their way in the originals or microfilm copies. The 1886 farm tax assessment lists are also being made machine-readable—with Optical Character Recognition—in a database that is made available on the Internet province by province.

The NHD issues typeset versions of the originals of the 1865, 1875 and 1900 censuses as well as the parish registers, easing access to the sources considerably. The alphabetical indexes included are of aid to both the professional and amateur researcher, since a person can be traced through them in a matter of seconds. For the largest towns microfiche replaces the books. In recent years it has become common to distribute the material in digital format, following a national standard for data entry and data distribution. The encoded versions of the censuses are mainly distributed on diskettes together with an optional statistical package.¹ In addition to other source material, both the 1865 and 1900 censuses are available for the whole country, including the parts registered by the Digital Archive.²

Jan Oldervoll of the Digital Archive is an expert on the use of these sources in history teaching, while Lars Nygaard has specialized in standardizing transcription formats. At the Norwegian Historical Data Centre, Terje Holtet is webmaster, Trygve Andersen is the computer specialist, Marianne Erikstad is the expert on the encoding of the family and occupation variables, and Gunnar Thorvaldsen is the expert on the use of the material in research, most notably on migration and mortality.³

¹ A prototype statistical interface and search engine are available on the Internet, at <http://www.rhd.uit.no/indexeng.htm>.

² Their website at <http://digitalarkivet.uib.no/index-eng.htm> contains these censuses as well as the 1801 census and a range of other source material.

³ Information on a number of researchers who have used the digitized Norwegian source material can be found in the general bibliography at <http://www.rhd.uit.no/art/littlst.htm> and the special bibliography for the censuses at <http://www.rhd.uit.no/census.htm>.

Publicly Available Nominative Censuses

Censuses based on manuscripts containing identifiable information such as names have been taken in Norway in 1801 and then with approximately ten-year intervals from 1865. The 1885 census includes only the towns and the eastern part of Finnmark province, the next complete census being taken in 1891. There is a special census for the towns from 1870, and several local censuses have been taken, especially for Oslo and Bergen. Also, the rhythm had to be changed because of World War II, when the census was moved from 1940 until 1946. The last complete census was taken in 1980. In 1990, sampling was used with varying density according to the size of each municipality. Norway's next— and probably, last — questionnaire-based census has been postponed until 2001. The plan is to collect household information from paper forms sent to all households, while using existing registers to assemble individual level data. In 1865 the *de jure* principle—just counting each individual where they had their formal residence was still used. In later censuses this was combined with the *de facto* principle, so that people would also be enumerated and reported, especially, where they were temporarily visiting or working.

Censuses up to and including 1900 can be used by anyone and for any purpose, except for the restriction imposed by the data protection registrar that individual level information about religion, ethnicity and illnesses should not be made public on the Internet. The Law on Statistics from 1907, however, stated that such material can only be used for statistical purposes, a clause that in 1986 was modified to “protect” the material for 100 years (Breivik 1988 #63). Thus, the 1910 census will not be made public until 2010, even though several researchers have been allowed, upon application to use this and later censuses for their statistical investigations, and even data process the material.⁴ Large parts of

⁴A summary overview of the information fields contained in these nominative censuses can be found on the web at www.rhd.uit.no/census/censustable.htm#Table2; for a similar overview of

the 1865, 1875 and 1900 censuses have been encoded for statistical purposes by the Norwegian Historical Data Centre (see Thorvaldsen 1995 #49) for an overview in English. Sivert Langholm has summarized social history research on the Oslo region with the 1865 to 1900 censuses as the main source material (Langholm 1974 #50; Langholm 1976 #51; the latter article is in English).⁵

1865

After half a century of statistical census taking only, the census taken in the beginning of January 1866 was a novelty in more than one way. Not only were names and other information about each inhabitant noted, for the first time the self-enumeration principle was introduced in Norway. In the towns, the house owners were requested to fill in the census forms themselves, forms which were to be distributed, collected, checked and if necessary completed by the district registrars. The forms should be signed by the enumerator, and for the town of Tromsø this field has been used to study the census takers. Only in two cases were the forms signed by the house owner, as a rule they were filled in and signed by some relative, some neighbor or someone else living in the house (Drake 1991).

This is in contrast to the report in the printed results from the census, where it is maintained that the house owner only had to be assisted in a few cases (Statistics Norway 1868-69). In the countryside, the parsons and the sheriffs assisted by the schoolteacher were still in charge of filling in the forms. At least formally, however, the ministers should direct the enumeration and control the census manuscripts before they were returned to the Department of the Interior. Complaints from the priests and the teacher about the increasing workload involved in census taking were not heeded by the authorities, but at least the teachers were paid to do census work (Tranberg 1994). Since they knew the

the more modern censuses, <http://www.rhd.uit.no/census/censnew.htm>.

⁵ The Norwegian censuses are well documented; instructions and sample questionnaires are available at <http://www.rhd.uit.no/census.htm>.

households better than other officials, the Ministry felt that their assistance could not be dispensed with. In order to cover the inhabitants of all farms, each enumerator received a copy of the printed farm tax assessment list for his district. These lists are organized topographically, and could be followed when going from farm to farm. The 1838 version of the tax lists was used for almost half a century until a new edition was printed. Since cottars' farms were not included in the lists and often are less centrally located, these were more likely to be under-enumerated. There are indications that the 1865 census is somewhat less complete than later censuses, but the main reason for this may be that it was the last to include only the *de jure* part of the population, so it was more likely to leave out temporary migrants.

The 1865 census was the first in Norway to record place of birth for each person listed. The parish was often noted instead of the municipality, and users should be warned that this may be the parish that a municipality belonged to at the time when the person was born, but which has later been changed (Thorvaldsen 1995, #31).⁶ For the last time, age was given instead of birth year, which may have led to errors because people reported their present age rather than the one they would reach in 1866 as instructed. As could be expected, round ages are over-represented, a phenomenon which occurs less frequently in later censuses (Thorvaldsen 1996). Family and household information should be given, but it has been found that in many cases a formal position such as "servant" was given instead of actual family relations.

Occupation was lumped into the same field as data on household status, and in addition to the preference for formal titles (*cottar*) rather than work functions, there is a widespread under-registration of secondary occupations. For instance the 1865 census is known to vastly understate the number of fishermen. In northern Norway and in some other places, the ethnicity of each individual should be specified, even if the forms provided no special field for this piece of information. In many cases this created problems, since some families were split and the census manuscript instead

⁶ See <http://www.rhd.uit.no/nhdc/summary.htm> for a summary of this dissertation which contains a number of source critical points concerning the 1865, 1875 and 1900 censuses.

organized according to ethnic belonging. The Histform standard requires that this information should be placed in its proper field, and this is the case for the part of the 1865 census entered by the Norwegian Historical Data Centre.

The original manuscripts are kept in bound volumes in the National Archives, with photocopies in some regional archives. The exception is the parish of Gol in Eastern Norway and some summary lists for Oslo which have disappeared.⁷ There are slightly different forms for towns and the countryside, the most notable being the address fields where farm number is left out in towns. Statistical results for the population are available in Statistics Norway 1868-69, while agricultural figures are reported in Statistics Norway 1869.

1870

This census covered only the towns, where the authorities felt they needed updated information because of rapid urbanization. From 1835 to 1920 the Province Governors (*Amtmenn*) sent five-annual reports to the statistical authorities. Late in 1870 they were ordered to include an urban census in their next report. The census should be taken and financed by the administrators of municipalities with formally recognized urban settlements (*byer eller ladesteder*). The Department of the Interior included a sample census form, but since these were to be printed by each municipality, the forms used are somewhat different from place to place (Isaksen 1988).

One town abstained from taking this census altogether, and the manuscripts for several others have gone missing, so the records for ten communities cannot be found in the 1870 collection in the National Archives. This material has not been microfilmed. Few if any of these census returns have ever been computerized or used in research, and it is not known if any statistical results were published.

⁷ The instructions to the census takers and sample census forms are available at <http://www.hist.uib.no/arkivverket/tellingdok.htm>.

1875

While the nominative basis of the 1865 census was a great leap forward, the methodological changes introduced in the census undertaken at the beginning of January 1876 meant a significant improvement in quality. This resulted both from the generally improving standards of education in Norway, and more specifically from the organizing of a special administrative agency for the production of statistics called *Det statistiske Centralbureau* (The Central Bureau for Statistics) in 1876. This had been prepared by employing more specialized personnel during the previous years. Most notable is the introduction of the explicit distinction between the *de jure* and the *de facto* population, with special fields for noting the whereabouts of absent people and the origins of temporary residents. However, the information about absent persons were noted on special lines at the bottom of each form, which sometimes makes it difficult to see which family or household they belong to, and creates extra problems when coding family relations. The standard specifications for computerized versions of Norwegian censuses, therefore, states that these persons are to be moved to where they belong among the persons present if possible (Nygaard 1995). Spot tests indicate that the 1875 census is even more complete than the previous one. For completeness and adherence to the *de jure* principle, census forms and instructions were also sent to Norwegian ships abroad via diplomatic channels, the forms being only marginally different from those used ashore. These despatches also contained letters with brief explanations in French and English.⁸

Another change was the reporting of birth year instead of age for each person. The intention behind this reform was probably to eliminate the confusion arising from the choice between reporting either the previous or the next birthday. Spot tests again indicate some improvement in correspondence with baptismal records, but since the census takers would ask for age and compute the year of birth, mis-reporting is still quite frequent although discrepancies of more than one year are seldom found. The

⁸ See: <http://www.rhd.uit.no/census/ships75.htm>.

improvement in the reporting of occupations was more significant, since the one field about both household and occupation was now split in two. Also, the census takers were explicitly told to report extra sources of income. Not only did this result in the mentioning of many adjunct jobs, it also became more customary to note the actual work function, and not only formal titles such as “cottar” or “lodger.” Even so the fishermen category is not complete, “primarily because grown-up sons living with their parents on the farms, and many servants who participated in the fisheries, where not noted as fishermen” (Solhaug 1976).

In the northernmost provinces, a special form was used with an extra column for ethnicity, while the margin still had to be used to note data on languages spoken. Instead of a general ethnic label, now the ethnicity of each of the parents should be noted. Therefore, we can easily measure the degree of out-marriage in ethnically mixed areas. Of course, when one of the parents was of mixed ethnic origin, the category “mixed” still had to be used. In essence, however, this reform probably led to more weight on the genetic rather than the cultural aspects of ethnicity. And even if spot tests indicate that the information on ethnicity as a rule is accurate, we should remember that a single person with different ethnicity from the rest of the household could easily be amalgamated. The information about languages spoken is often rich, indicating that many people were trilingual, speaking Sami and Finnish, as well as Norwegian. But we ought to be suspicious of information where ethnicity is not in accordance with language. For instance, if the form reads that Sami people spoke Norwegian, it probably means that they spoke Norwegian in addition to Sami (Thorvaldsen 1996).

The main results from this census are reported in Statistics Norway 1882.⁹ The census manuscripts from major parts of the country have been computerized by the Norwegian Historical Data Center, most notably Oslo, Bergen, Trondheim and northern Norway. Now a 2 percent household sample of households from the rest of Norway is being digitized.¹⁰

⁹ See <http://www.rhd.uit.no/census/ft1875.htm> for a summary of the proceedings in Norwegian.

¹⁰ For an overview of municipalities covered, see RHD 1997 or the web at: <http://www.rhd.uit.no/datakilder.html>.

1885

Also in 1885 the census was intended to cover only the towns where rapid urbanization had led to the need for an update since the next national census had been postponed until 1891 in order to be synchronized with the censuses taken in many other countries. (After the establishment of the special Statistical Bureau, Norway participated in the international statistical conferences). Unlike in 1870, census forms were financed and printed by the Statistical Bureau, while local authorities were responsible for the carrying out the enumeration, including its costs. However, the scope of this census was extended to cover five rural municipalities in the easternmost part of the northernmost province Finnmark, districts close to the border with Russia. This was allegedly because the authorities wanted to map the large Finnish population, which was (without reason) viewed as a security risk because of their ties with Finland, ruled as semi-independent grand-duchy under Russia until 1917.

The information found is similar to that given in 1875, but in addition there is a field with information about employers. Only the manuscripts from a few towns covered in the 1885 census have been computerized, but there is a printed publication giving aggregate statistics (Statistics Norway 1887). As for the 1891 and 1900 censuses, spot tests indicate that a number of manuscripts have been lost.

1891

With a growing number of inhabitants and variables to handle, the census authorities were on the look-out for ways to streamline the process of taking a census. While the 1865 census forms were continuous and the later censuses used one form for each domicile, the 1891 census for the first time introduced one sheet for each individual. Since the forms can be sorted by the contents of each variable, this probably speeded up the enumeration process. Later on the archives had to resort the material according to topography. However, it is a hassle to copy all the individual sheets and the accompanying ones for each

domicile, so this census is the only one up until the 20th century which was never microfilmed. Certain parts, like Bergen, have been copied, but otherwise the lack of backup copies has hitherto blocked the computerization of the 1891 census.

This is a problem since the original is rather inaccessible, both because it is archived centrally in the National Archives, and because going through all the sheets requires a lot of page-turning. So far, it has not been discovered that any significant parts of the census have gone missing because of extra work involved in arranging this kind of material. However, the agricultural part of the census was destroyed in a fire in 1939, so only statistical results on farm production are available.

From 1891 on, the forms used in towns contained special fields in order to note information about each flat in apartment buildings, and the number of variables about domiciles was increased considerably. This source may, therefore, be called the first combined census of people *and habitation*.

On the personal forms, two new variables were introduced. First, in addition to a field specifying the person's own occupation, an extra field gave the occupation of the primary breadwinner in the household. This makes it easier both to be more specific about occupations and relationships within the family or household. Also, an appendix to the instructions lists occupational titles that required further specification. For instance the census takers were ordered to note what kind of work a mechanic did, and where he was employed.¹¹

Second, it was to be noted in a separate field whether or not the two spouses were related by blood ties, probably a concession to researchers interested in issues of heritability. In several of the fields, the census takers could underline keywords or abbreviations instead of or in addition to writing an ordinary string of letters. For instance in the marital status field there were preprinted terms like 'Unmarried' and 'Widowed' and in the occupational field the words "Self-employed" or "Employed" should be underlined. The person's own ethnicity should be given, and there was a special field for language in the parts of Norway that are ethnically mixed.

¹¹ A similar list instructed them to add the name of the province to any ambiguous municipality name.

The statistical results from the 1891 census are available in several volumes (see Statistics Norway 1894-98).

1900

This was the first Norwegian census to be tabulated mechanically, using the kind of equipment with punched cards and electro-mechanic counters that Herman Hollerith had developed for the US census of 1890. In addition to saving work at the time, this meant the reintroduction of forms for each domicile and the dropping of the individual sheets since there was no longer any need to sort individuals manually. So far, the 1900 census is the most recent to have been made public (in 1960) and has also been microfilmed by the Church of Latter-Day Saints.

Even if the layout of the forms is very different from 1891, the contents of the fields are quite similar. In both censuses there is a special form for information about each apartment in the towns. There are no questions for additional information, except for the specification of birth date for children two years old or younger. On the other hand, the questions about the occupation of the primary breadwinner and blood ties between spouses were dropped. Agricultural information was taken, but has survived for the towns only, since the remainder burnt in 1939. In the relevant districts, the forms contained extra fields, one about ethnicity and one about language. When the forms were processed in the Central Bureau of Statistics, however, ethnicity information was crossed out and corrected if not in harmony with data on the person's language. Thus, a cultural definition of ethnicity prevails in the statistical results. As a rule the ethnicity data originally noted is legible and has been transferred into the machine readable version which exists for the whole country.

The 1900 census is available in machine readable formats for the whole country from the Digital Archive and Norwegian Historical Data Centre—that is with the exception of a number of lists which cannot be found in the regional archives (mainly for urban areas). The statistical results were published in six volumes plus a “Main Overview” (*Hovedoversigt*) also containing the instructions to the census takers and copies of the forms used (Statistics Norway 1906).

Digitizing the Source Material

An early preoccupation with northern Norway led to the completion of the electronic versions of the 1865, 1875 and 1900 censuses for the northernmost provinces. In addition, the censuses for several municipalities have been transcribed, for example, the largest cities in southern Norway. For 1865 and 1875, this material complements the significant parts covered by the Digital Archive, so that the whole country is covered. The NHD maintains a database with information on historical, nominative sources that have been made machine readable in Norway. This is available both on the Internet and as a biannual publication with explanations in English. The transcription of parish registers is a more labor-consuming undertaking than the censuses, making geographic coverage more limited. The handwriting in these sources is often Gothic, increasing the transcription difficulties significantly.

The first stage in the treatment consists of a direct transcript of originals or xeroxed copies from the Riksarkiv (The National Archive) or from the several Statsarkiv (Regional Archives). The personnel perform a word - by - word transcript, using PCs. After the transcripts undergo proof-reading, any errors found are corrected. For extra safety, spot tests are carried out, and if the result is unsatisfactory, repeated proof-reading will be undertaken. The guiding principle is to be true to the source material. Most information is written down exactly as it appears, with a few minor exceptions in order to enhance the information value and userfriendliness of the material. Even so, all users must be aware that the resulting database will contain errors and inconsistencies, most of which also can be found in the original sources.

Usually, each type of source for a *kommune* (municipality) or *prestegjeld* (parish) is treated on a one-by-one basis, but it is possible to assemble two or more of these localities into a more comprehensive geographical unit before sorting on whatever key is desired. Since all records are related to addresses, it is possible to construct consistent geographical entities that may be compared over time. Both municipalities and the higher administrative unit of the *fylke* (province) have been made comparable in this manner.

The machine-readable censuses exist in a verbatim full text transcript as well as an encoded format. To the latter end, the next step is to standardize variables with a view to statistical treatment. The numeric codes for occupation, family status and county of birth, were originally made to standardise the information in the sources for the production of statistics. The codes are, however, also useful in the process of record linkage, because the information on family status and birthplace can be employed in a uniform way. A semi-automatic encoding program enables operators to complement the textual variables with numeric codes, creating input files for statistical software packages. Again, consistency across time and space is stressed so that results from cross-sectional analysis of several censuses may be comparable. (Thorvaldsen 1995-I). Personal names and farm names are not coded or standardised, so inconsistencies in name spelling must be dealt with by researchers who want to identify people in several sources. Agricultural data on livestock and sowing exist in the 1865 and 1875 censuses, and can be linked to the information about persons. Users are warned that there is an arbitrary element in all coding procedures, so researchers ought to check the coded version against the original.

Why Censuses Should be Encoded

The ultimate aim of the coding process is to compute statistics on the basis of the information about the citizens counted in the censuses. But why cannot social historians and demographers make do with the printed tabulations that were published shortly after each census was taken?

Even though the original tables will be valuable for some researchers, there are at least three reasons why new statistics must be computed for most purposes. First, the printed tables were made with a definite geographic area, typically a parish or a province, in mind. Often the boundaries between the parishes changed in the period between two censuses, making comparisons over time quite difficult. Second, this problem of comparison is aggravated by the fact that the system for categorizing the information in the censuses was revised each time a new census was taken. Occupational groupings are not consistent across consecutive censuses. These are

serious limitations, since the historian will always want to study the development of society.

Even so, many researchers will consider the third reason to be the most important. The printed statistics at hand utilize only some of the variables in the censuses. And moreover, these variables are crosstabulated in a very limited number of ways. For example, you will find that occupations are counted for each sex, but not distributed according to age groupings or civil status. The publications from the Bureau of Statistics may tell you that in some parishes there were many farmers and also that the elderly made up a large proportion of the population. If you conclude that most farmers were aging, you are at risk of committing an ecological fallacy, since you only combine the variables at the parish level. To test this kind of hypothesis, the variables studied should be combined at the individual level.¹² This can be done most efficiently by entering the census into the computer, subsequently encoding the variables of interest and finally treating them with a statistical program package.

The Procedure of Encoding

Such software typically will be able to deal with numeric information only. Therefore the textual information contained in the original census must be translated into digits. At the same time the information is standardized, so that irrelevant shades of meaning are eliminated. Whether the craftsman is called a joiner or a carpenter, he should be given the same occupational code. Technically speaking, this can be done in a number of ways, and I now turn to a discussion of some of them.

The first computer projects in the field of social history emulated the social scientist and coded the information manually before it was put into the computer. Since computer storage was expensive, space could be saved by just noting numerical codes instead of the full text from the sources onto forms before transferring these numerics to Hollerith cards. The drawbacks of

¹² See Sivert Langholm. (1976). "On the Scope of Micro History," *Scandinavian Journal of History*, page 13ff.

this procedure were, fortunately, quickly realized. Once some cross-tabulations had been made, the researchers found that the numerical categories did not quite suit the aim of the research. Then the whole procedure of encoding and punching had to be started from scratch. This was also necessary if another scholar wanted to use the same material to test other hypotheses. Since historians like to stay in close touch with their source material, it was agreed that a much more flexible procedure is to start by entering the complete content of the source material into the computer. Then sorting algorithms can be used to produce lists that give an overview of the content of the census, and programs can be written that partly automate the coding.

Much work can be saved by using software that has been specially designed for the purpose of coding censuses and other highly structured sources. The use of dedicated software will also ensure that the coding is done in a consistent manner. One must remember that the same concepts will be repeated over and over again in a census. There are many persons who are called "farmer" in the column of occupations. Their occupations can be coded simultaneously with one command. Also, the same concept may be written in several ways that are slightly different. You should, therefore, be able to specify that words are truncated to the left or to the right, or use mask characters within a word.¹³

Since the Norwegian universities had bought mainframes from different computer companies, three distinctive coding systems were developed to go with the different operating systems used on these computers. Such a program system for coding nominal sources will consist of a suite of programs, that must be run in a sequence to produce the desired. The first step is to split the content of the sources, keeping the values of each single variable together in separate files. This program also sorts the file and compresses it so that each value on that variable is repeated only once. Together with the textual value there will be room for the numeric code and also how many times that variable is represented in the source being coded. Each line starts with a

¹³ A mask character represents several letters, for example, it is common to let the question mark represent any vowel, consonant or digit.

sequence number that points to a special file (pointer file) which contains references to the source file, so that the codes can be put into the right records when the coding job has been done.

Such compressed files must be constructed for each variable that cannot be coded automatically, normally *position in the household, occupation and place of birth*. Straightforward information such as *sex* and *age* is mostly translated directly into numerics without any operator intervention.

Now the proper work of encoding can start, and for the sake of efficiency and standardization several methods should be combined. If a similar source has been coded for another district, one should start with precoding. This implies that values which are equal in two districts will get the same codes. For each district where the censuses are encoded, the codes are transferred to a database, and over time this collection of codes has speeded up the coding work significantly.

However, additional values will turn up in every new source that is treated. Also, some codes cannot be copied uncritically from one district to the next. For instance "carpenter" will often signify a sailor along the coast, while he built houses in the interior. Earlier, such instances were coded with a command driven system where you could specify that all values containing a certain string of letters should get a specific code (automatic coding). You could also key in a specific code on line number X (manual coding). With the introduction of interactive, full screen oriented software, this has been replaced by interactive coding. Now, a sequence of variable values will be marked using the mouse or arrow keys and then given the proper code. Since the text of the variable is sorted, this has turned out to be quite efficient.¹⁴

Each time a census has been coded, it should be thoroughly controlled either on the screen or a printout. When checked and corrected, the codes can subsequently be transferred to the numeric file, which is the end-product of the encoding process. A special program will combine the compressed file and the pointer file so that the sequence of the original source file can be reconstructed. Of course, only the numeric codes and not the value texts are

¹⁴ A procedure for using MS Access to do the encoding, is described in the methodology book *Databehandling for historikere* by Gunnar Thorvaldsen.

transferred. In a second step codes for the straightforward variables, for example sex and age, are written to the numeric file. If there are any problems at this stage with, for instance, sexless persons or people who are 150 years old, this is dealt with interactively. As a result, we have a standard ASCII file where each person is represented by a record of numerics. This file can be read into standard statistical program packages such as SPSS or NSDstat¹⁵, given that its contents are specified in a file with variable and value declarations.

Encoding Complex Variables

The simple variables are given numerical codes like 1 for the unmarried, 2 for the married, etc. Not much more need be said about this, except that it is wise to reserve a special code like zero or a multiple of 9 for instances where information is lacking. Other variables are, however, far from simple to code in this respect, since they can be interpreted in several ways and are used by researchers for a number of different purposes. Social scientists and historians have used much time to develop intricate systems for the encoding of these variables, and I intend therefore to go into more detail about three of these variables.

Place of Birth

Mostly a census will specify a person's birthplace to the level of the municipality or parish, which in Norway will signify the same territory.¹⁶ For instance, in connection with the study of migration, it is necessary to systematize the places where people lived and where they were born. This is done by assigning a code of four digits to each parish in Norway. All parishes in the same

¹⁵ NSDstat is a simple, menu driven statistical package, also available as a Windows version. It was developed by the Norwegian Social Science Data Services. (see the review in *History and Computing*, 1992.)

¹⁶ A Norwegian "parish" will as a rule will be bigger and more populous than the English variant, but smaller than an English county. The abbreviation "prgj", which you find in Figure 12-2 stands for *prestegjeld*, which means parish. The Norwegian "province", which consists of several parishes or municipalities, will mostly be bigger than a county.

province have the first two digits in common, thus making it easy to study migration both at the level of the parish and the province on the basis of the same codes. When the third digit is a zero, it means that the municipality in question is a town. For example 1902 is the town of Tromsø in the province of Troms, 1929 is the parish of Berg in the same province, while 2012 is the parish of Alta in the neighbouring province of Finnmark.

The number of municipalities has changed over time. Parishes are for numerous reasons split, joined together, and areas are transferred from one municipality to the other. To keep track of this, the numbering system is dynamic. Each area that has ever been a parish in its own right, has been given a unique code, and there is a special database which works as an inventory of all the changes. This also contains the various names that have been associated with a given municipality throughout the last two centuries. To deal with immigrants, the system can easily be extended with country codes for the rest of the world.

Even if the coding system is dynamic, extensive and well documented, some problems arise in actual use. Quite a few parish names are ambiguous in the sense that several municipalities are called the same. If the two alternatives are far apart, this ambiguity can be resolved by choosing the code for the place that is closest to where the census was taken. Where this criterion is not effective, the code for unknown must be used. This is also the case when only the name for a smaller area, such as the name of a farm is given, since many place names can be found in several parishes. Sometimes the name of a territory that comprises several parishes is given. If the whole territory lies within a province the code for that province can be given, leaving two zeros for the parish part of the code. 1900 would thus mean somewhere in the province of Troms.

Occupations

While place names are coded as a hierarchy with two levels, occupations are coded along two distinct dimensions. One dimension is the trade, that is to say the type of work a person did or the kind of business he or she was employed with. The other dimension is the place that the person's job had in the social hierarchy at the time when the census was taken. The classification

of occupations has been based on the system developed for the 1875 census by the Central Bureau of Statistics, but has been extended by several researchers. Each of the two dimensions is normally given a two-digit code.

The first digit will grossly designate if the occupation belonged to the primary, secondary or tertiary type of trade. For instance, farmers and fishermen are given codes 11 and 13 respectively, while artisans and those employed in industry will be given 21 and 24. Some occupational groupings, such as the artisans, are very heterogeneous, and some researchers will therefore need to specify such trades further with two more digits. 21 04 could then mean a plumber, whereas 21 11 would mean a carpenter. It is therefore right to say that the classification of trades can be done in a hierarchy with three distinct levels. The most detailed level is, however, not used by many researchers.

Let us now turn to the other dimension, the social hierarchy. Sociologists have constructed a number of methods for the ranking of social groups. Most of them are, however, unsuitable for historic research, since they presuppose modern tools such as the opinion poll. Instead, we have built on nineteenth century criteria, for instance the ownership of land or the formal ranking of artisans. These social groupings, though, do not make up one consistent hierarchy, but rather several lateral hierarchies that cannot easily be compared. For example, in the countryside people are classified according to whether they own land, rent land or are simply cotters. Artisans are grouped as masters, journeymen or apprentices. Civil servants are ranked from those that were royally appointed down to the ordinary clerk.

By using two dimensions, most occupations can be classified in a detailed way, without making the system very complex with a lot of categories. But unfortunately, the censuses quite often do not provide us with the information we need to encode a job according to both dimensions. A "worker" can be given a place in the hierarchy, but his trade cannot be resolved. A joiner's trade is easy to decide, but his ranking must be stated as missing in the encoded file. Some phrases are very ambiguous. Is the "fisherwife" rowing the boat or is she merely the wife of a fisherman? Such dubious cases can sometimes be interpreted by looking at the household to which the person belonged. On the whole, information about the

occupations of married women is scarce in nineteenth century censuses. In northern Norway farming and fishing was a usual combination, and normally the wife ran the farm while her husband was out fishing. The census, however, will simply state that the latter was a peasant and a fisherman, saying nothing about the wife. Therefore, the occupational code of the head of the family should be copied to a special variable for the rest of the family members. Then the historian will be able to deduce the kind of work that women and children did.

Position in the Family and Household

Since there can be several families within one household, ordinarily position in the family and in the household are encoded as two distinct variables. It is, for example, common for a person to be at the same time a son in the family and a servant to the household. The grouping of individuals as wives, sons, other kin, lodgers and servants is mostly straightforward and can be done by looking at the compressed file only. Time and again, though, it will be necessary to inspect the whole household to distinguish between the family and the household variable.

Such mechanisms become a prerequisite if also the relationships between persons belonging to the same family or household is to be represented in the encoded file. The serious student of family and household structure need not only count how many sons, but also whose sons they were. For that purpose the family or household code of the father or mother must be placed in a separate variable in each son's record. This obviously cannot be done by using the compressed file, since there all relationships between persons have been removed. The process of coding these relationships, therefore, to a great extent must be done manually, on the basis of information in the original source file. Advanced software that will semi-automate this process is being developed, and we hope this will make much more census material available for advanced family research.

Studying Migration in Troms Province

One project has used much of this information for a detailed

study of migration with individual level data in an entire province (Thorvaldsen 1995-II). This treatise analysed the migration that took place in the province of Troms in northern Norway during the second half of the nineteenth century, the period when net in-migration turned into net out-migration from this region. With the censuses as my main source material, I estimate the amount of migration both in and out of the province, between its constituent communes and within the communes. Then I describe to what extent the in- and out-migrants differ from other people with reference to characteristics like sex, civil status, occupation, etc. My empirical aim is to explain the relative amount of migration compared to other parts of Norway, and what made the different types of migrants move, while other people stayed put in the same place. The project was financed by the Norwegian Research Council.

Much effort had to be put into the development and description of new methodology, much of it based on the extensive use of computers. The research project is the first one entirely dedicated to the study of historical migration in Norway. It is also the first attempt to study the development of an entire province ("fylke") with statistics based on individual level data. In this connection, out-migrants from the municipalities are traced to their new domicile. Moreover, for the first time in Scandinavia, programs for automatic record linkage were used in historical research.

The empirical results are summarized in a new migration model based on the concept of the frontier. In nineteenth century Norway, internal migration can be studied in the ministerial records and in the censuses. The sources are similar to the Danish ones, and we do not have the longitudinal catechismal church records they have in Sweden. From 1865 on emigrants to destinations overseas were registered in special ledgers by the police. These are much more complete than the migrant records kept by the local priests from 1812 on as part of the ministerial records, where at least half the migrants are missing from the lists. Several studies show that the migrant records are biased with respect to social class, distance travelled and period studied. These records, therefore, are very far from a reliable basis for the production of statistics about nineteenth century migration. This

leaves us with the censuses.

We can divide the methods applied into three groups. First I use the published statistics from the Central Statistical Bureau to compare the level of migration in Troms with other parts of Norway. Second I do a cross-sectional study of different groups of migrants by cross tabulating the variables found in the censuses. To be able to compare the communes and the variables over time, this must be based on a renewed, computerized treatment of the census returns. Thirdly, I do automatic record linkage on the census population for three communes in order to follow individual migrants and other residents in more than one source. The second and third group of methods then, constitute the main part of my thesis, and here I combine information about the population of Troms at the individual level.

Troms had one of the highest population growth rates in Norway during the second half of the nineteenth century. How does the province compare with other parts of Norway when it comes to the amount of migration that was going on? Since we have gross migration figures only for emigration overseas, nothing but statistics on net migration can be calculated for domestic migration. There can, however, be no doubt that in terms of direct effects, the rapid growth was due more to the surplus of births over deaths than to in-migration. In fact, net in-migration to Troms only lasted until about 1880.¹⁶ In relative terms, the national out-migration of 5 percent was heavier than that of the northernmost provinces (1 to 3 percent), partly explaining the higher population growth in Northern Norway. On the other hand, also the birth surplus of 15 percent contributed, compared to the national average of 12 percent. If one takes the indirect effects of in-migration on fertility into consideration, however, there can be no doubt that in-migration of young, fertile people was a crucial factor behind the high population growth in Northern Norway.

Emigration is an important part of this picture, Troms being the Norwegian province with the smallest relative number of emigrants (2 percent versus 6 percent for Norway). When it comes to the relative number of in-migrants in the communes, Troms

¹⁶ Between the censuses of 1875 and 1900 net out-migration reduced the population size by 5000 persons.

matched the national level of 23 percent in 1865. In 1910, however, the national percentage had risen to 34, while Troms stayed at a virtually stable 24 percent. Simultaneously the proportion of in-migrants declined from 13 percent to 9 percent. In 1910 only 1 percent of the population of Troms was born in southeast Norway.

This surprised me, since migration from this region to Troms was essential during the first six decades of the 19th century. Over time, not only did in-migration decline in relative terms, but also the in-migrants were recruited to a greater degree from northern Norway. In 1865, 44 percent of the in-migrants to the Troms communes were born in Southern Norway and abroad; at the turn of the century this number was only 27 percent. The explanation behind this is two-fold—first an expansion of the regional potential migrant pool because of the growing population in northern Norway and second the reduction of the potential migrant pool among South Norwegians who more frequently chose to migrate to the U.S. Thus, reduced in-migration is a more important explanation than emigration behind the positive net out-migration figures for Troms after 1875.

The direction of the migration flows changed over time in other ways as well. Twenty-five percent of out-migrants from the communes in Troms were found in Southern Norway or were counted as emigrants in 1865. In 1900 this figure had more than doubled (54 percent). According to the 1865 and 1875 censuses that we have computerized, the number of migrants born in Troms who stayed up north in Finnmark by far outnumbered those who had left for the more southern parts. In the 1900 census there were more migrants from Troms who had migrated to the south than who lived in Finnmark.

Thus the flow of migration out of Troms not only increased in size during the period under study, its main orientation had shifted as well. New opportunities for employment in the construction sector in the province of Nordland were the main cause of this development. At the level of the province, Hägerstrand's theory (1947) of the constant migration fields does not fit very well, neither for in- nor for out-migration. It also strikes me that so few people from Troms settled in Lofoten, the area of the important cod fisheries that so many of them visited each year. Therefore, it seems that the migration area only partly overlaps

with the area of the population's social network, which is the part that gives advantageous employment opportunities. I also found a rural-urban barrier in that people from Tromsø tended to settle in towns once they left the province. To some extent, this may be related to an urban social network, but a more important reason behind the barrier probably was the special skills needed in many urban occupations.

It is part of Tilly's model (1977) that circular migration over time often develops into chain migration. This is the probable mechanism behind the fact that so many migrants from Troms settled in Finnmark - many of them and their relations had participated in the annual fisheries. Also chain migration of people with general skills to Finnmark was widespread, even if information about the opportunities there was common property. Social relations had an important role to play when migrants settled in a new place, not only as a channel for information. In what ways and to what extent did the different types of migrants differ from the resident or non-moving part of the population?

Here I primarily base my results on cross-sectional analysis of the three censuses, but I shall also use findings from the attempts at record linkage. A computer program divided the population of each commune into residents, in-migrants and out-migrants. The latter group makes up some 9 percent to 10 percent of the population in each census, while the proportion of in-migrants declined from 13 percent to 9 percent during the period studied. Among the in-migrants the men were in a small and decreasing majority throughout the period, whereas the women among the internal migrants formed a growing majority (56 percent to 59 percent). (They formed a clear majority also among the migrants inside the communes.) For the out-migrants from the province, however, the picture is more complicated. Here women dominated in 1865 and 1875, whereas men dominated in 1900. This was not because a higher proportion moved the long distance to Southern Norway during the last decades. Rather, most men went to Finnmark to partake in the rich fisheries, while Southern Norway attracted female servants. Thus Ravenstein's law that women only dominate the flows of short-distance migration does not seem to hold true. My conclusion is that the differences in the migration of the sexes to a great deal depended on sexual differences in the

labour market.

The occupation as a fisherman could provide a decent income for an independent young man. For a bread-winning father, however, it was an insecure position, which reduced the family income since it was harder to put the hands of the wife and the children to work. Also fishing was a tough way of life when old age approached - the proportion of fishermen was correspondingly lower in the highest age groups. For these reasons it was very desirable for the young man to find other means of subsistence for his family than fishing only. In my opinion two strategies were the most likely during this period.

First he could try to become a farmer by inheriting, buying or finding a tenancy. Most tenant farms were, however, sold out early in this period. The Norwegian laws of allodium make buying a farm an insecure option as long as any inheritor is still alive. Therefore, inheritance comes out as the prime option, and we have seen that the peasants in Northern Norway had adopted the custom of splitting the farm among the children. It is probable that the child who moved away to other communes or even provinces, by doing that would risk his customary right to inherit his share of the land. Thus the importance of the fishing/farming sector in the economy of Northern Norway reduced the amount of migration substantially, both through the inheritance factor and its flexible capacity to employ any number of hands.

The second strategy was to find employment in the growing "modern" sector of the economy with the trades, small factories, plants, commerce and other services. When choosing this strategy, migration became a tempting alternative. For one thing most of these jobs were in the growing towns and small urban centres. And the possibility that a suitable position would turn up would increase if people took a wide geographical area into consideration. From this perspective it was only natural that many migrants in and from Troms are found in the said sector of the economy. As the in-migrants were qualified for so many of these jobs, there were consequently fewer openings in this sector for the natives and less incentive for them to migrate. A large proportion of the Finns was servants according to the 1865 and 1875 censuses. Their diminishing numbers towards the end of the century, therefore, were not due only to assimilation, but also to the precarious

position of the servant group when the rapidly growing number of children took over their tasks.

There is much truth in the saying that people mostly migrate to conserve their old way of life. For the fishing peasants of Troms it was difficult to keep up this combination of trades if they moved to town. Agriculture was marginal on the coast of Finnmark, there was little room for new farms to the south and small chances to fish on the American prairie. In my view the ways the population of Troms migrated and the reasons why they stayed, were to a high degree related to the economy, here represented by the four frontiers. Peasants tended to stay put, while people in the trades and the services tended to migrate. The social network only comes in as a secondary determinant, helping people decide exactly where to go.¹⁷

Conclusion

One example of research on the Norwegian material must suffice here, but there are others in the fields of mortality studies, family history, name research and the study of emigration. While the three completely digitized censuses give good opportunities for diachronic and regional comparison within Norway, the infrastructure for bilateral or multilateral international comparisons of historical processes and social structures is much less well developed. The initiatives taken within the IMAG group and by the IPUMS-International project described elsewhere in this volume, make probably a more efficient bringing together of the many interesting—and sometimes incompatible - research findings from many countries.

¹⁷ A more comprehensive summary of the dissertation can be found at url <http://www.rhd.uit.no/nhdc/summary.htm>

