The SweCens NAPP Variable Documentation for the Swedish censuses 1880, 1890,1900 and 1910

Table of contents

1   Introduction

This documentation has been developed within the Swedish project "Encoding and linking Swedish Censuses" (SweCens). The project has been financed by two grants from the Swedish Research Council and by funding from the NAPP project at Minnesota Population Center (MPC) at the University of Minnesota, USA.

SweCens I

The SweCens I project (2010-2011) was headed by Per Axelsson at the Centre for Sami Research (CeSam) at Umeå University. The project was a joint effort including participants from Umeå University, the University of Gothenburg, Lund University, the Swedish National Archives and the Stockholm City Archives. The purpose of the project was to encode the Swedish 1890 census according to NAPP principles and to produce an updated encoding of the 1900 census. The project also evaluated methods for record linkage between the 1890 and 1900 censuses and developed methods for linking the censuses to the POPUM database at the Demographic Database at Umeå University. There were also a number of research topics within the project. These topics could be analyzed using the encoded information and the linked data. Another goal of the SweCens I project was to prepare a new application for funding of an even greater project that would include registration, encoding and linking of all Swedish censuses from the first census in 1860 and onwards.

Participants in the SweCens I project
Per Axelsson (Phd), Centre for Sami Research, Umeå University
Mats Berggren, IT Dept, Swedish National Archives.
Martin Dribe (Professor), Dept of Economic History, Lund University.
Sören Edvinsson (Associate Professor), Demographic Database, Umeå University.
Johan Gidlöf, Stockholm City Archives.
Maria Larsson, Demographic Database, Umeå University.
Christer Lundh (Professor), Dept of Economic History, University of Gothenburg.
Carl Szabad, SVAR Dept, Swedish National Archives.
Maria Wisselgren (Phd), Demographic Database, Umeå University.

SweCens II

The SweCens II project (2013-2016) has received funding from the Swedish Research Council and the main purpose is to digitize and encode the Swedish 1930 census and make it available for international research. The project is headed by Anders Nordström, head of the SVAR Dept at the Swedish National Archives. The reference group for the project is the same group as participated in the earlier SweCens I project (see above). As of March 2016 the Swedish 1930 census has not been completely digitized. Currently the aim is to have the census complete and encoded in 2018. The digitization of the 1930 census has taken longer time than originally expected. One reason for this is that the digitization of the 1910 census took longer time than expected. It was decided within the project to finish the 1910 census before transferring all resources to the 1930 census.

2   NAPP editions of the Swedish censuses

There have been several editions of the Swedish censuses over the years.

The 2008 edition of the 1900 census
In April 2008 the project Sweden1900 produced a first NAPP encoded version of the Swedish 1900 census and a test version of the 1890 census containing only one county. The work was funded by the NAPP project at the MPC. The encoding in 2008 was done by Mats Berggren, Johan Gidlöf and Mats Hayen (Phd), Stockholm City Archives. The ethnicity variable was coded by Per Axelsson.

The 2011 edition of the 1890 and 1900 censuses
The 1890 census was encoded within the SweCens I project (see above). The 1900 census was also updated and the two censuses were delivered to the MPC in November 2011. The encoding in 2011 was done by Mats Berggren, Johan Gidlöf and Carl Szabad. The ethnicity variable was coded by Per Axelsson. The 2011 edition contained the following changes and improvements compared to 2008:

- There was a major revision of the coding of occupations and other work variables. The coding was made consistent between the 1890 and 1900 censuses.

- In 2011 occupations was coded with three code variables according to NAPP-HISCO and three additional code variables according to the HISCO 2002 standard. The variable LABFORCE was also encoded.

- In 2008 only the most common family relation phrases were encoded. In 2011 all family relation phrases in the 1890 and 1900 censuses were examined and encoded if possible.

- The software used to construct the RELATE and RELATEI variables was improved.

- The encoding of disabilities was more thorough. Three variables were produced giving information both of the main disability and also about additional disabilities.

- Additional geographic encoding was made in order to give more accurate residence and birthplace variables.

- A number of counter variables that were produced by the Swedish staff in 2008 were omitted in the 2011 edition. These variables were: FAMSIZE, NFAMS, PRMFAMSZ, GQ, NCHILD, NCHLT5, ELDCH, YNGCH, NSIBS, NCHLT10, MARRYDAU, MARRYSON, UNMARDAU, UNMARSON, UNMARKID, NCOUPLES, NMOTHERS, and NFATHERS.

- A number of variables referring to Swedish storage media for scanned census images were no longer relevant and were omitted in the 2011 edition.

The 2012 edition of the 1880, 1890 and 1900 censuses
The 1880 census was completed and NAPP encoded in 2012. The encoding in 2012 was done by Mats Berggren, Johan Gidlöf and Carl Szabad. The ethnicity variable was coded by Isabelle Brännlund (Phd) at the Centre for Sami Research, Umeå University. The editions of the 1880, 1890 and 1900 censuses that were produced and delivered in November 2012 contained the following changes and improvements:

- The 1880 census was delivered complete and new.

- For the 1890 and 1900 censuses the occupation variables were updated with more occupations encoded.

- Ethnicity was encoded and delivered for all three censuses.

The 2016 edition of the 1910 census
The digitization of the 1910 census was completed in September 2015 and the NAPP encoding was completed in March 2016. The NAPP encoding was done by Mats Berggren, Johan Gidlöf and Carl Szabad (who sadly passed away in October 2015). The ethnicity variable was coded by Isabelle Brännlund. The same methods and code lists were used for the 1910 census in 2016 as for the 1880, 1890 and 1900 censuses in 2012. There was only one minor difference compared to earlier censuses:

- In 1910 the occupation titles for "servants" changed compared to earlier censuses. This has affected both the occupation encoding and the encoding of the variables RELATE and RELATEI.

3 The Swedish census database FOLK

The digitized Swedish censuses for the years from 1860 to 1930 are stored in the database FOLK which is administered by the department SVAR within the Swedish National Archives. Currently the censuses 1880, 1890, 1900 and 1910 are complete while the census 1930 is under work. There are also a few counties registered for the census years 1860 and 1870. Nothing is registered for 1920. The database FOLK is an MSSQLServer database. The database is available for searching over the Internet. A specially developed web application gives users access to the database.

The most important table within the database FOLK is the table FOLK which contains a record for each person. Here is a short description of the fields in the table. The fields in italic are fields that have been added to the table after registration. The other fields are the fields that have been produced by the registration software. Every field is described with its name, its MSSQLServer datatype and a short descriptive text.

Table 1: The Swedish National Archives census database FOLK

Field:	Datatype:	Description:
ID	*Int*	*Unique numeric identifier(temporary)*
FRNR	Smallint	Census year (1880, 1890 or 1900)
LANNR	Smallint	Countynumber (1..25)
FORSNR	Smallint	Parishnumber within county
PNR	Int	Personindex within parish
ABNR	Int	Alternative identification code for parish
ARKBILD	varchar(50)	Parishname
KONTRAKT	varchar(50)	Contract, part of a diocese
SCBKOD	varchar(9)	Geographical code
LAN	varchar(50)	County name
SVARNR	varchar(6)	Identifier for microfiche volume
KORTNR	Smallint	Microfiche within microfiche volume
SIDA	varchar(6)	Page in source
RAD	Smallint	Row in source
HEMORT	varchar(70)	Place of residence
HNR	Int	Householdindex within parish
FAMST	varchar(70)	Position in family as transcribed
FNR	Int	Family unit within household
FAMSTKOD	varchar(1)	Code for position in family
FORNAMN	varchar(70)	First name
ENAMN	varchar(70)	Last name
TITEL	varchar(70)	Title as transcribed
YRKE	varchar(70)	Occupation as transcribed
YRKKOD	varchar(30)	Swedish occupation codes at the time of the censuses
STAM	*Varchar(70)*	*Unique numeric identifier (stable over time)*
LYTE	varchar(70)	Disability as transcribed
FODAR	varchar(10)	Birth year
FODORT	varchar(50)	Birthresidence as transcribed

FODFORS	varchar(75)	Birthparish as transcribed
FSCBKOD	varchar(9)	Code for birthparish
KON	varchar(1)	Sex
CIV	varchar(1)	Marital status
TROSB	varchar(70)	Religion as transcribed
UTKYRK	varchar(1)	Person has exited the Swedish church (1900-field)
UTKYRKT	varchar(70)	Text about the exit from the church (1900-field)
NATIONAL	varchar(70)	Nationality
FRANVARA	varchar(70)	Absent
KOPPOR	varchar(1)	Vaccination for small pox
EJDOP	varchar(1)	Person not baptized in the church (1900-field)
STATUS	varchar(1)	Status of the record
KLAMMER	varchar(1)	Existence of household marker in the source
STRUKEN	varchar(1)	Person crossed over in source
MANTKOD	varchar(1)	Parish/rote of residence is the persons taxation parish
ROTE	Int	Part of the City of Stockholm
MANTORT	varchar(50)	Taxation parish if other than parish/rote of residence
KYRKORT	varchar(50)	Church parish if other than parish/rote of residence
KALLANM	varchar(78)	Remark about the source
OVRIGT	varchar(255)	Additional information
ANM	varchar(255)	General remark
RPLATS	varchar(5)	Registration place (Workplace)
RSIGN	varchar(4)	Registration signature (User)
REGDAT	varchar(8)	Registration date (Timestamp)
FODD	varchar(4)	Birth year
BILDNR	varchar(10)	Picturenumber on CD
CDNR	Smallint	CD-number
CDNAMN	varchar(15)	CD-name
FODLAN	varchar(50)	County of birth
FODLANKOD	varchar(9)	Code for County of birth
XFORNAMN	*char(4)*	*For SOUNDEX-search of first name*
XENAMN	*char(4)*	*For SOUNDEX-search of last name*

Source: Database FOLK, Swedish National Archives

The field ID is a unique identifier that can change over time when the database is updated. The field STAM is a unique identifier that is stable over time and can be used to link to the NAPP encoded data. The table also has yet another alternate unique key consisting of the fields FRNR, LANNR, FORSNR and PNR. These four fields together form a unique key for a record in the table FOLK. Ordering the table by these fields gives the registration order of the records.

This paper describes how NAPP variables are created from the Swedish 1880, 1890, 1900 and 1910 censuses. The information in the NAPP variables is derived from the information stored in various fields in the FOLK table in the FOLK database.

4   The Swedish NAPP variables

This section contains a list of all NAPP variables that will be produced from the Swedish 1880, 1890, 1900 and 1910 censuses. Five variables, CNTYHISTSE, OCCHISCO, OCCMULTISE, DISABCO2 and DISABCO3 are new in the 2011 edition compared to 2008. See the code book document: ***SweCens_Codebook.pdf***

Table 2: Swedish NAPP Household variables

Field	Field Name	Type
Technical Variables (Household)		
Record type **(H)**	RECTYPE	varchar(1)
Country of residence	CNTRY	varchar(3)
Census year	YEAR	varchar(4)
Number of people in household	NUMPERHH	Integer
Residence name as transcribed	RESNAME	varchar(70)
Household database-ID	HHIDSE	Integer
Householdindex within parish	HNRSE	Integer
Geography Variables (Household)		
Urban/rural status	URBAN	varchar(1)
Countycode, stable over time	COUNTYSE	varchar(2)
Parishcode, stable over time	PARSE	varchar(9)
Countycode, historical county code at the time of the census	CNTYHISTSE	varchar(2)
Parishindex within county at the time of the census	PARNRSE	Integer

Source: Swedish National Archives

Table 3: Swedish NAPP Person variables

Field	Field Name	Remark
Technical Variables (Person)		
Record type **(P)**	RECTYPEP	varchar(1)
Person index within household	PERNUM	Integer
Household database-ID	PHIDSE	Integer
Person database-ID	PIDSE	Integer
Person index within parish	PNRSE	Integer
Person crossed over in source	CROVERSE	varchar(1)
Person not present	NOTPRESE	varchar(1)
Constructed Family Interrelationship Variables (Person)		
Mothers location in household	MOMLOCSE	Integer
Probable step/adopted mother	STEPMOMSE	varchar(1)
Rule for linking mother	MOMRULESE	varchar(1)
Fathers location in household	POPLOCSE	Integer
Probable step/adopted father	STEPPOPSE	varchar(1)
Rule for linking father	POPRULESE	varchar(1)
Spouses location in household	SPLOCSE	Integer
Rule for linking spouse	SPRULESE	varchar(1)
Location of grandmother (side unidentified) in household	GRANDMOMSE	varchar(4)
Location of grandfather (side unidentified) in household	GRANDPOPSE	varchar(4)
Location of mother's mother in household	MOMMOMSE	varchar(4)
Location of mother's father in household	MOMPOPSE	varchar(4)
Location of father's mother in household	POPMOMSE	varchar(4)
Location of father's father in household	POPPOPSE	varchar(4)
Potential step grandmother (ambiguous side)	STEPGMOMSE	varchar(1)
Potential step grandfather (ambiguous side)	STEPGPOPSE	varchar(1)

Field	Field Name	Remark
Position in family as transcribed	FAMINFSE	varchar(70)
Family unit within household	FAMNUMSE	Integer
Family unit membership	FAMUNITSE	Integer
Code for position in family	FAMPOSSE	varchar(1)
Code for family relations	FAMRELSE	varchar(4)
Code for family relations reference	FAMREFSE	varchar(1)
Code for birth status	FAMBIRSE	varchar(1)
Rule for setting RELATE and RELATEI variables	RELRULSE	varchar(1)
Demographic Variables (Person)		
Relationship to household head	RELATE	varchar(4)
Age	AGE	varchar(3)
Sex	SEX	varchar(1)
Marital status	MARST	varchar(1)
Birth year	BIRTHYR	varchar(4)
Relationship to household head, IPUMS-international codes	RELATEI	varchar(4)
Information about absent person as transcribed	ABSENTSE	varchar(70)
Nativity and Birthplace Variables (Person)		
Country of birth	BPLCNTRY	varchar(5)
Born in NAPP country	NAPPSTER	varchar(2)
Code for county of birth, stable over time	BPLSE	varchar(2)
Code for birthparish, stable over time	BPLPARSE	varchar(9)
Nationality	NATNALTY	varchar(3)
Birthparish as transcribed	BPARSE	varchar(75)
Birthresidence as transcribed	BRESSE	varchar(70)
Ethnicity and Language Variables (Person)		
Ethnic origin first stated	ORIGIN	varchar(4)
Religion, first stated	RELIGION	varchar(4)
Religion as transcribed	RELIGSE	varchar(70)
Work Variables (Person)		
Labour force participation	LABFORCE	varchar(1)
Occupation, NAPP-HISCO-code	OCCHISCO	varchar(5)
Occupation as transcribed	OCCSTRNG	varchar(70)
Relationship information in occupation string	OCRELATE	Integer
Status information in occupation string	OCSTATUS	Integer
Swedish occupation codes at the time of the census	OCCSE	varchar(30)
Occupation, HISCO-code, 2002 standard	HISCOSE	varchar(5)
Relation code, HISCO 2002 standard	HISCRELSE	Integer
Status code, HISCO 2002 standard	HISCSTATSE	Integer
Flag for multiple occupations	OCCMULTISE	varchar(1)
Disability Variables (Person)		
Disability as transcribed	DISABSE	varchar(70)
Disability code	DISABCO	varchar(2)
Disability code, second	DISABCO2	varchar(2)
Disability code, third	DISABCO3	varchar(2)
Other Variables (Person)		
Title as transcribed	TITLESE	varchar(70)
Last name	NAMELAST	varchar(70)
First name	NAMEFRST	varchar(70)
Person not baptized in the Swedish church	NOBAPTSE	varchar(1)
Person has exited the Swedish church	CHEXITSE	varchar(1)

Source: Swedish National Archives

5   The Swedish NAPP household variables

This section describes how the NAPP household variables are created for the Swedish 1880, 1890, 1900 and 1910 censuses.

Technical Variables (Household)

The variable **RECTYPE** is set to "H" for household. The variable **CNTRY** is set to the NAPP-code for Sweden. The variable **YEAR** is set to the census year. This variable corresponds to the field FOLK.FRNR in the FOLK database. The variable **NUMPERHH** contains the number of people in the household. The variable **RESNAME** contains the text from the field FOLK.HEMORT. The field contains place of residence as transcribed. The variable **HHIDSE** contains the unique numeric value from the field FOLK.STAM for the person that is the head of the household. This facilitates linking between household records and person records and linking to the original FOLK-table. To further ensure backward compatibility between the NAPP variables and the FOLK-database the Swedish household number (FOLK.HNR) is kept in the variable **HNRSE.**

Geographic Variables (Household)

The variable **URBAN** contains urban/rural status for a household. In the Swedish 1880, 1890, 1900 and 1910 censuses every parish was described as either being rural or being part of a city. The coding of the variable URBAN is based on this division of the Swedish parishes. It must be noted that the division isn't very exact. A "city-parish" could contain vast rural areas and a "rural-parish" could contain a growing suburb of a neighbouring city. It is however the best way to categorize urbanity because it will be consistent with the Swedish population statistics of the period.

These values are used in the variable **URBAN.**

Code:	Explanation:	Comment:
1	Rural	
2	Urban.	
9	N/A: Enumerated on board ship	Not applicable in Sweden.

The variable **COUNTYSE** contains a two digit county code that is stable over time. The code range from 01 to 25 but the value 02 is excluded. The two digits correspond to the first two digits in the 9-digit parish code in the variable **PARSE.** This parish code is stable over time and used both in the 1890 and 1900 censuses. This code can be used to link people between census years.

The variable **CNTYHISTSE** contains the county code used at the time of the census. For 1880, 1890, 1900 and 1910 the code values range from 01 to 25. Before 1968 the city of Stockholm had the value 01 and the county of Stockholm had the value 02. In 1968 the city of Stockholm became a part of the county of Stockholm and this new county got the number 01 and the value 02 was excluded.

The variable **PARNRSE** gets its values from the field FOLK.FORSNR in the FOLK database. This is the parish index within county from the Swedish 1880, 1890, 1900 and 1910 censuses. The field is important for backward compatibility.

6   The Swedish NAPP person variables

This section describes how the NAPP person variables are created for the Swedish 1880, 1890, 1900 and 1910 censuses.

Technical Variables (Person)
The variable **RECTYPEP** is set to "P" for person. The variable **PERNUM** is a person index within a household. The variable **PHIDSE** contains the unique numeric value from the field FOLK.STAM for the person that is the household head of the household. The variable serves as a unique household identifier connecting persons to households (see above under Technical Variables (Household)). The variable **PIDSE** is set to the unique numeric value in FOLK.STAM for the current person. The variable **PNRSE** is set to the value in the field FOLK.PNR. This variable is needed for linking to the FOLK database.

In the registration forms, that are the source for the Swedish 1880, 1890, 1900 and 1910 censuses, some persons are crossed over. This could for example be done for people that also were registered in another parish. A person was crossed over in the form for the parish where he or she didn't belong. In the FOLK-table all persons are registered but the persons that have been crossed over in the source are marked with a flag in the field FOLK.STRUKEN. All people in the FOLK-table are converted to NAPP-records and a variable is created in order to mark persons that have been crossed over in the source. The variable **CROVERSE** has the following simple value list:

Code: Explanation:
0   Not crossed over in the source
1   Crossed over in the source

In the Swedish Census some people in each parish were listed as "obefintliga". The word corresponds to "not present" or "not existing" in English. It means that the person had been living in the parish but wasn't known to be present at the time when the census was taken. A phrase containing the word "obefintliga" was normally recorded in the place name field in the census. Usually the list of "not present" would appear last for each parish after the inhabitants in the last village or hamlet had been listed. A variable **NOTPRESE** is added to indicate if a person is listed as "not present". The value list is as follows:

Code: Explanation:
0   Not listed as "not present".
1   Listed as "not present" (Swedish: obefintlig).

Constructed Family Interrelationship Variables (Person)
It isn't trivial to create the family variables and demographic variables in NAPP from the Swedish 1880, 1890, 1900 and 1910 censuses. The basic problem is that the Swedish way of encoding family relationships differ from the other NAPP censuses. The Swedish censuses lack a code for "relation-to-household-head" that is directly comparable to the codes in other countries. The system for family relations in The Swedish census has to be described in detail.

Family relations in the Swedish 1880, 1890, 1900 and 1910 censuses
The family relations in the Swedish census database FOLK have been coded using a very simple method primarily aimed to record the relations within nucleus family units. The method used for recording relations is as follows: Within a household every person is given a family unit number, stored in the field FOLK.FNR. In the field FOLK.FAMSTKOD every person is given a simple code that describes the person's position within the family unit. The codes are:

Code:	Explanation:
F	Far (Father)
M	Mor (Mother)
B	Barn (Child)
E	Ensamstående (Solitaire)

With this simple system it is easy to establish relations within a family unit but impossible to establish, with any certainty, relations between two persons belonging to different family units within the same household. Simple family households can be completely described but it isn't possible with extended family households or multiple family households where there are two family units within the same household. There is no way to establish with certainty that a person coded as "father" in family unit 2 also is "child" to the "father in family unit 1.

The field FNR does NOT correspond exactly to the NAPP variable **FAMUNITSE.** The variable **FAMUNITSE** should have the same value for all related people within the same household. All members of an extended family with a father and/or mother, a married child and a grandchild will have the same value in **FAMUNITSE.** In FOLK this isn't the case. The members of the extended family would be divided in to two "families" were the father and/or mother would have FNR=1 and the married child and its child would have FNR=2. This means that the variable **FAMUNITSE** has to be constructed in some way for the Swedish 1880, 1890, 1900 and 1910 censuses in order to be compatible with the same variable in the other NAPP censuses. The Swedish family number field FNR is added as a variable with the name **FAMNUMSE**.

The source also contains some additional information about relations that are recorded as text in the field FOLK.FAMST. This information is directly transcribed from the source. The information can be "mother in law", "his mother", "her stepmother", "her child in an earlier marriage" and so on. This information can be the basis of a more detailed coding of relations within family units and households. The textfield FAMST is added to NAPP as a variable **FAMINFSE.** The field FOLK.FAMSTKOD for position within family is added as a variable **FAMPOSSE.**

Enhancing the description of family relations in the Swedish censuses
It was decided in 2007, before the coding of the 2008 edition of the Swedish 1900 census, to evaluate the possibilities to encode the information found in the field FOLK.FAMST. The purpose would be to make the Swedish 1880, 1890, 1900 and 1910 censuses more compatible with the other NAPP censuses. Coding the field FAMST might make it possible to establish stepparent relations and multi

generation relationships in the Swedish census. The contents in the field FAMST in the database FOLK have been analyzed. For every unique field value the number of instances of that value has been counted. The result is presented in the following table.

Table 4: Number of instances of unique values in the field FAMST in database FOLK
The first column is a classification based on the number of instances per value. One instance is a personrecord where the field FAMST has the value. The second column is the number of unique values in the class. The third column is the total number of instances within the class. The fourth column is percent of total instances per class.

Instances per value	Number of values	Number of instances	Percent
>= 100000	3	2997404	87.53
10000 – 99999	9	119908	3.50
1000 – 9999	48	143979	4.20
100 – 999	297	85984	2.51
10 – 99	1418	39561	1.15
2 – 9	6170	20886	0.61
1	16482	16842	0.48
Total	24787	3424564	100.00

Source: Database FOLK, The Swedish 1900 Census, Swedish National Archives

As can be seen from the table a very limited number of values are extremely frequent and a large number of values only occur a few times. This makes encoding possible and meaningful. An encoding of the 1775 values in the first five classes (values occurring 10 times or more) would give a hit ratio of 98.89% which is a fairly good result. In 2011, 2012 and 2016 more than 98% of the relation phrases occurring in the 1880, 1890, 1900 and 1910 censuses have been encoded. Remaining phrases have been impossible to encode and have been given the value 9997 for "unknown".

Encoding family relations in FAMST
An attempt has been made to define an encoding schema for the information in the field FOLK.FAMST. In this attempt three variables are used. The first variable **FAMRELSE** is a code for the family relation expressions found in the field FAMST in the database FOLK. The majority of the codes are identical to the codes used in the NAPP variable **RELATE**. Codes marked with an asterisk (*) are those NOT found in the codelist for the NAPP variable **RELATE**.

The codes in the variable **FAMRELSE** does NOT necessarily refer to a relation to the household head. Instead the values in the field FAMST usually refer back to some preceding adult person, a man or a woman, within the household, but NOT necesserely the household head. In a household with many generations the phrase "hennes oä. d." ("her illegitimate daughter") can be found in FAMST on a child. That child is perhaps the child of a woman who has the value "d." in FAMST

which means that she herself is a daughter, probably of the household head and his wife.

The values in FAMST often, but not always, contain a possessive pronoun like "hans" and "hennes" ("his" and "her") or a noun in genitive like "mannens" and "hustruns" ("the man's" and "the wife's") together with a word describing a family relation. The second variable **FAMREFSE** is a code that describes to which person or persons the relation expression is referring. Relation references can be coded with codes for "his" (the man's or the husband's) and "her" (the woman's or the wife's). This makes it easier to establish what person within the household that is referred. It also makes it possible to establish stepfather/stepmother-relations. There is a common code for the phrases "deras" ("their in common") and "bådas" ("belonging to both"). The code for "unspecified" is used for those cases where a relation is described without any further explanation. The field FAMST may contain only the phrase "s." ("son") . Then it may be assumed that it refers to the person's father and mother. If the phrase is "syster" ("sister") it probably refers only to the household head.

Common values are "Enkan" and "Enkl." ("Widow" and "Widower"). They have been coded with a special code (9995). The information is already coded in the NAPP variable **CIV**. The only purpose with a special code for "Widow/widower" in **FAMRELSE** is to compare the result with the recorded values in **CIV**. All records that have blanks or a NULL-value in the field FAMST will get the value 9999 "Missing" in the variable **FAMRELSE**. Persons getting 9999 are either household heads or solitaires, for example domestic servants unrelated to other household members.

The third variable is a code for "birthstatus" **FAMBIRSE**. The field FAMST often contains information about illegitimacy or a specification of which marriage a person was born in. The phrases can be "hans s. i 1:a g" ("his son in his first marriage") or "hustruns d. f. ä." ("wife's daughter before marriage") or "hennes oä. d." ("her illegitimate daughter"). These distinctions are kept by setting codes for "born within marriage", "born within an earlier marriage", "born before marriage" and "illegitimate child". It would be easier to code only "child" but then the distinctions used in the source would be lost. Researchers may for example be interested in the number of illegitimate children within different social groups.

All three variables **FAMRELSE, FAMREFSE** and **FAMBIRSE** have to be set to produce a complete encoding for a person record. The encoding is based on the principle that the values should be coded directly from the list of unique values in FAMST. There should be no need to view the records in the database to decide what code to use. That would be equivalent to doing the census registration all over again. The following three tables give the complete code lists:

Table 5: Family relation code for the variable FAMRELSE
The first column gives a family relation expression in text and the second column contain the code

Relationship:	Code:
Wife	0201
Child	0301
Adopted child	0302
Stepchild	0303
Fosterchild	0306
Child-in-law	0401
Step child-in-law	0402
Parent	0501
Stepparent	0502
Parent-in-law	0601
Stepparent-in-law	0602
Sibling	0701
Step/half/adopted sibling	0702
Sibling-in-law	0801
Grandchild n.s.	0901
Step grandchild	0903
Grandchild on son's side	0905 *
Grandchild on daughter's side	0906 *
Other relatives n.s.	1001
Grandparent fathers side	1014 *
Grandparent mothers side	1015 *
Aunt/uncle	1021
Nephew, niece	1031
Cousin	1041
Great grandchild	1051
Widow/Widower (a very common unclassifiable phrase)	9995 *
Unknown (not yet encoded)	9997
Missing (no value in FAMST)	9999

Source: The Swedish National Archives

Table 6: Family reference codes for the variable FAMREFSE
The first column gives a family reference type in text and the second column contain the code

Relationship:	Code:
Unspecified	1
Their (in common) or 'belonging to both'	3
His (father's, the mail's)	4
Her (mother's, the woman's)	5
Unknown (not yet encoded)	8
Missing (no value in FAMST)	9

Source: The Swedish National Archives

Table 7: Birth status for the variable FAMBIRSE
The first column gives a birth status in text and the second column contain the code

Relationship:	Code:
Unspecified (normally within the current marriage)	1
Within an earlier marriage	3
Born before marriage	4
Illegitimate	5
Unknown (not yet encoded)	8
Missing (no value in FAMST)	9

Source: The Swedish National Archives

The following table shows an attempt to code the 60 most common values in the field FAMST using the code lists presented earlier. The 60 first values are those values that have more then 1000 instances each. Together they represent 95.23% of all records that have a value in FAMST.

Table 8: Values in field FAMST coded with new relation code
The first column gives the values in the field FAMST. The second column gives the number of instances. The third column is the new relation code applied to the corresponding FAMST values.

FAMST value	Number of instances	FAMRELSE	FAMREFSE	FAMBIRSE
s.	1122137	0301	1	1
d.	1063484	0301	1	1
h.	811783	0201	1	1
h:u	15399	0201	1	1
Syster	14559	0701	1	1
oä. s.	14478	0301	1	5
bådas s.	14300	0301	3	1
oä. d.	14244	0301	1	5
bådas d.	13714	0301	3	1
Moder	11765	0501	1	1
Barn	11182	0301	1	1
hennes oä. s.	10267	0301	5	5
hennes oä. d.	9848	0301	5	5
fosterd.	7729	0306	1	1
b. s.	6192	0301	3	1
b.	6076	0301	1	1
Fader	5899	0501	1	1
Broder	5887	0701	1	1
b. d.	5855	0301	3	1
fosters.	5750	0306	1	1
Fosterson	5347	0306	1	1
Mor	4344	0501	1	1
dennas oä. s.	4279	0301	5	5
dennas oä. d.	4132	0301	5	5

Fosterdotter	3903	0306	1	1
d:r	3862	0301	1	1
Måg	3601	0401	1	1
hans s. i 1:a g.	3188	0301	4	3
h. oä. s.	3021	0301	5	5
Son	3020	0301	1	1
h. oä. d.	2903	0301	5	5
Far	2857	0501	1	1
Svärmor	2651	0601	1	1
Svärmoder	2648	0601	1	1
hans d. i 1:a g.	2618	0301	4	3
s:n	2530	0301	1	1
Dotter	2164	0301	1	1
m. s. i 1:a g.	2160	0301	4	3
Svägerska	1980	0801	1	1
Fosterbarn	1949	0306	1	1
Bror	1836	0701	1	1
m. d. i 1:a g.	1728	0301	4	3
Enkan	1666	9995	1	1
deras s.	1597	0301	3	1
Svärfar	1583	0601	1	1
hennes d.	1563	0301	5	1
Mågen	1513	0401	1	1
deras d.	1444	0301	3	1
hennes s.	1424	0301	5	1
Dotterson	1388	0906	1	1
hu.	1358	0201	1	1
Enkl.	1325	9995	1	1
fosterb.	1251	0306	1	1
syst.	1232	0701	1	1
Enka	1190	9995	1	1
Svärfader	1145	0601	1	1
hennes s. i 1:a g.	1134	0301	5	3
h. h.	1130	0201	4	1
hennes d. i 1:a g.	1062	0301	5	3
hustruns s. f. ä.	1017	0301	5	4

Source: Database FOLK, the Swedish 1900 Census, Swedish National Archives

It is possible to encode the information in the field FOLK.FAMST in to the new NAPP variables **FAMRELSE, FAMREFSE** and **FAMBIRSE.** It will enhance the quality of the NAPP family variables and demographic variables for the Swedish censuses. The encoding work is reasonable if only the more frequent values are encoded. In the 2008 edition of the 1900 census the values that occurred at least 10 times were encoded. In the 2011, 2012 and 2016 editions of the 1880, 1890, 1900 and 1910 censuses all family relation strings have been examined and encoded if possible.

Using the new codes to establish NAPP family and demographic variables
The phrases in the field FAMST have been coded and based on that encoding the other variables have been given values. The values in the other family and demographic variables have been set by programming. It must be discussed with MPC if this is a fruitful method or if the family and demographic variables should be set in some other way. We are open to discuss all possibilities. The encoding of the variables has to be evaluated by the MPC.

The codes in **FAMRELSE, FAMREFSE** and **FAMBIRSE** are used together with the values in the fields **FAMNUMSE, FAMPOSSE, CIV** and **SEX** to produce more accurate values for the NAPP family variables and demographic variables. A child's **MOMLOCSE** and **POPLOCSE** can be set using the **FAMPOSSE** variable. The variables **FAMRELSE** and **FAMREFSE** provide the extra information needed to determine if a child's mother or father is in fact a stepmother/stepfather. This makes it possible to produce the variables **STEPMOMSE** and **STEPPOPSE** with reasonable accuracy.

The **MOMRULESE** and **POPRULESE** variables are given values stating what rule has been used in setting the **MOMLOCSE** and **POPLOCSE** variables. The current value lists for the **MOMRULESE** and **POPRULESE** variables are not applicable in the Swedish case. The **MOMLOCSE** and **POPLOCSE** variables in the current version of the Swedish 1900 Census are not constructed from relation variables (**RELATE** or **RELATEI**). Instead the variables are constructed from the Swedish fields FOLK.FNR and FOLK.FAMSTKOD as has been explained earlier. We have used a new code value in the present version of the Swedish 1900 census. The code list used is:

Code:	Explanation:
0	No father (mother) of this person present in the household.
5	For the Swedish Census the linking to father/mother is based on the values in the Swedish family relations variables.

The variables **SPLOCSE** and **SPRULESE** can be correctly created for almost every person because spouses always are members of the same family unit. The present value list for **SPRULESE** is not really appropriate for Swedish circumstances. We have used a new code value in the current version of the Swedish 1900 Census. The code list used is:

Code:	Explanation:
0	No spouse of this person present in household.
9	For the Swedish Census the linking to spouse is based on the values in the Swedish family relations variables.

The variables **FAMRELSE** and **FAMREFSE** together with **FAMNUMSE** and **FAMPOSSE** make it possible to establish multi generation relations and to give correct values to the variables for location of grandparents and stepgrandparents. The table below shows how persons with the value "s." in FAMST have been coded in the field FOLK.FAMSTKOD.

Table 9: Values in the field FAMSTKOD for the value "s." in FAMST

The value "s." is the most common value for "son". The first column gives the values in the field FAMSTKOD. The second column gives the number of instances.

Value in FAMSTKOD	Number of Instances
M (mother)	8
F (father)	8046
E (solitaire)	513
B (child)	1113570

Source: Database FOLK, The Swedish 1900 Census, Swedish National Archives

The table reveals that 8 sons have been coded as mothers. This is probably some kind of registration error. It also shows that 8046 sons ("s." in FAMST) have been coded as fathers ("F" in FAMSTKOD). This indicates that they have their own families but are children to the parents in another family unit within the same household. It is possible to establish **POPLOCSE** and **MOMLOCSE** relations for these persons. By doing so it is also possible to give values to the variables that describe multi-generation relationships. These variables are **GRANDMOMSE, GRANDPOPSE, MOMMOMSE, MOMPOPSE, POPMOMSE, POPPOPSE, STEPGMOMSE** and **STEPGPOPSE**. The variable **FAMUNITSE** can, with some programming effort, be created from the Swedish 1880, 1890, 1900 and 1910 censuses. As explained earlier the variable **FAMUNITSE** does NOT correspond directly to the Swedish variable **FAMNUMSE.**

Demographic Variables (Person)

The variable **AGE** can easily be calculated from the Swedish fields FOLK.FRNR (Census year) and FOLK.FODAR (Birth year). The variable for marital status **MARST** can be given values based on the values in the corresponding Swedish field FOLK.CIV. The conversion table will be as follows:

Code:	Explanation:	Code in FOLK:
1	Married, spouse present	G (Gift)
2	Married, spouse absent	G (Gift) (See text below).
3	Separated	Not applicable in Sweden
4	Divorced	X (Frånskild)
5	Widowed	E (Änka/Änkling)
6	Never married/single	O (Ogift), S (Sammanb.), F (Förlovad)
7	Indeterminate/unknown	U (Okänd)
8	Illegible	Not applicable in Sweden
9	Blank, missing	If the field is blank, corrupt or NULL

The table below shows the frequency of the values used in the field for marital status in the Swedish 1900 census database:

Table 10: Frequency of values in the field CIV in the database FOLK

Value in FOLK	Number of Instances
O	3185399
G	1693675
E	314150
X	6023
U	812
F	52
S	0

Source: Database FOLK, The Swedish 1900 Census, Swedish National Archives

As can be seen from the table the value "F" (Förlovad = Engaged) is used only 52 times and the value "S" (Sammanboende = Living together) isn't used at all. The value "2=Married, spouse absent" in **MARST** requires some consideration. In the Swedish data there is a field FOLK.FRANVARA which contains information about people absent during the census. Their marital status is however not affected by this in the Swedish database. They are coded as "married" although they can be recorded absent with a phrase like "Working in Norway". The value "2=Married, spouse absent" is set for all women who have husbands with a text in the field FRANVARA.

The variable **SEX** corresponds to the Swedish field FOLK.KON. The conversion table will be as follows:

Code:	Explanation:	Code in FOLK:
1	Male	M (Man)
2	Female	K (Kvinna)
8	Illegible	O (Okänd)
9	Missing/blank	If the field is blank, corrupt or NULL

There is no value for "unknown sex" so the Swedish value "O" for okänd=unknown has to be converted to 8=Illegible.

The variables **RELATE** and **RELATEI** contain codes that describe how a persons relation to the household head. As described earlier the Swedish 1880, 1890, 1900 and 1910 censuses do not have a corresponding field. The variables **RELATE** and **RELATEI** have to be given values based on the values in the Swedish family relations fields described earlier. It requires some consideration but it is possible to use the variables **FAMRELSE, FAMREFSE, FAMNUMSE** and **FAMPOSSE,** in combination to produce values for the variables **RELATE** and **RELATEI.** Relations within the same family unit can be established with accuracy, but it is more difficult establishing the right values in **RELATE** and **RELATEI** for persons not belonging to the same family unit as the household head. It can however be done for a fair number of people. The table below shows an example of how the **RELATE** variable can be constructed from the other variables.

Table 11: Family relationships to household head

namelast	sex	marst	Faminfse	famnumse	famposse	*Famrelse*	*Famrefse*	*fambirse*	***relate***
Gran	2	5	Enka	1	M	*9995*	*1*	*1*	***0101***
Gran	2	6	d.	1	B	*0301*	*1*	*1*	***0301***
Thorsell	1	1	Måg	2	F	*0401*	*1*	*1*	***0401***
Thorsell	2	1	h.	2	M	*0201*	*1*	*1*	***0301***
Thorsell	1	6	s.	2	B	*0301*	*1*	*1*	***0901***

Source: Database FOLK, The Swedish 1900 Census, Swedish National Archives

The first six variables in the table correspond directly to fields found in the Swedish database FOLK. The next three fields are the code fields that encode the information found in the variable **FAMINFSE** that correspond to the family relations text field FAMST in FOLK. With the use of the first nine variables the tenth variable **RELATE** can be given values by a program that implements a set of rules and assumptions. One rule is that the first person in the household is defined as the household head. It is assumed that the person coded as "son-in-law" in the second family must be son-in-law to the household head in the first family. It is also assumed that the wife to the son-in-law must be a daughter to the household head. A third assumption is that if the parents in the second family are children/children-in-law to the household head then the children in the second family must be grandchildren to the household head.

Most values in the variables **RELATE** and **RELATEI** describe family relations but there are also values for "non-relatives" such as servants and lodgers etc. For the Swedish censuses 1880, 1890 and 1900 these values are set by a simple rule. If a person that is a "non-relative" has the value "5." in the variable **OCCSE** (the Swedish occupation code at the time of the census) than the person gets the **RELATE** code 1211 (Servant) otherwise the person is coded as 1230 (Other non-relative). In 1910 there was a change in the use of occupation titles for servants. Some designations became much more frequent compared to the earlier censuses. In the 1910 census persons are coded as 1211 if they have the values "5." or "5.1" in the variable **OCCSE**. Relatives of persons coded as servants (1211) always get the code 1219 (Relative of employee) in all censuses. It must be noted that the original Swedish occupation coding (the variable **OCCSE**) wasn't always very exact and there may be somewhat too few people coded as 1211 (Servant) in the censuses 1880, 1890 and 1900 and somewhat too many in the census 1910.

In the Swedish 1880, 1890, 1900 and 1910 censuses we have added a variable **RELRULSE** that gives information about how the values in the variables **RELATE** and **RELATEI** have been set. The variable **RELRULSE** has the following value list:

Code:	Explanation:
0	The values in RELATE and RELATEI are based directly on information in other variables in the persons record.
1	The values in RELATE and RELATEI are derived from information in other persons records. Typically this is done for people in a second family unit within a household. A child in a second family gets the RELATE value based on its parents relation to the household head in the first family unit in the household.

The variable **BIRTHYR** (Birth year) can be created easily. It has a corresponding Swedish field FOLK.FODAR. The Swedish field FOLK.FRANVARA contain information about persons that are absent during the census. This information can be added as a NAPP variable **ABSENTSE**. The text in the variable will be in Swedish, and not understandable by all researchers, but the mere existence of a value in the field indicates that the person was absent during the census.

Setting family and demographic variables
The family variables and demographic variables in the Swedish 1880, 1890, 1900 and 1910 censuses are generated and set in five steps.

First step: The family variables **FAMINFSE, FAMNUMSE** and **FAMPOSSE** and the demographic variables **AGE, MARST, SEX, BIRTHYR** and **ABSENTSE** are set directly from the corresponding variables in the database FOLK. The household variable **NUMPERHH** is also set directly based on the number of household members in FOLK.

Second step: Based on the encoding method described earlier the variables **FAMRELSE, FAMREFSE** and **FAMBIRSE** are given codes based on the information in the variable **FAMINFSE** (the variable FAMST in FOLK).

Third step: The father, mother and spouse pointer variables **MOMLOCSE, MOMRULESE, POPLOCSE, POPRULESE, SPLOCSE** and **SPRULESE** are set based on the information in the fields **HNRSE, FAMNUMSE** and **FAMPOSSE.**

Fourth step: On the basis of the information in the variables already set it is possible to give values to another set of variables. In this step **STEPMOMSE, STEPPOPSE, GRANDMOMSE, GRANDPOPSE, MOMMOMSE, MOMPOPSE, POPMOMSE, POPPOPSE, SPEPGMOMSE, STEPGPOPSE, RELATE, RELATEI** and **RELRULSE** are set.

Fifth step: When the family and demographic variables have been given values it is possible to set the value to one final variable: **FAMUNITSE.**

The variables set in step one and two are reliable in the sense that they derive from the source and are not set by an application. The variables in step three to five are set by a rather complicated application. If MPC would like to produce the family variables and demographic variables in some other way it would be safest to start with the variables set in step one and two. They all emanate directly from the source and are not manipulated by an application. The problem with this approach is that there exists no **RELATE** variable in the Swedish Census. The **RELATE** variable is produced by the Swedish software in step four. It is set using the values in other variables, among them are **MOMLOCSE** and **POPLOCSE**. If MPC will use the **RELATE** variable as a basis for calculating other variables this fact has to be considered. We have made an attempt to give the family variables and demographic variables values using our specially developed application. It is up to MPC to decide how the information should be used. The MPC can code some of these variables using their own applications.

Using surnames to establish family relations

Names are NOT used for coding family relations. It is theoretically possible but there are a number of issues that has to be taken into consideration. Most important is the transition in the Swedish use of surnames that takes place during the period during which the 1880, 1890, 1900 and 1910 censuses is taken. Before the middle of the 19th century the use of patronymic names was prevailing. At the end of the 19th century there was a shift towards surnames. The children would then always get the surname of their father and not a patronymic name. The names being used as surnames were in most cases old patronymic names. The son of "Anders Eriksson" would call himself "Oskar Eriksson" instead of "Oskar Andersson" as he would have done a generation earlier. In the Swedish 1880, 1890, 1900 and 1910 censuses it can be difficult to determine if the name registered in the surname field is actually a real surname or a patronymic name. The principal system for making patronymic names was to add "son" or "dotter" to the father's first name. In the database the surnames are registered in the way they were written in the source. A common practice was to abbreviate patronymic names for daughters by replacing "dotter" with "dr" or "d:r". For example the name "Andersdotter" was often written as "Andersdr". There are also a few special cases for patronymic names for sons. If a fathers name was "Erik" his son's patronymic name could be written either as "Eriksson" or as "Ersson". If a fathers name was "Olof" his son could be called "Olofsson" or "Olsson".

Another consequence of the transition in the naming system was that wives began to use their husband's surnames after marriage. When patronymic names were used women would always keep their patronymic names after marriage.

It must also be considered that most children do not have any information in the surname field in the Swedish census database. This can be true even for adult children. An adult daughter living with her parents may have an illegitimate child but no surname.

Nativity and Birthplace Variables (Person)

An important part in the NAPP coding of the Swedish 1880, 1890, 1900 and 1910 censuses is the coding of geographical information. An effort is made to enhance the coding that has been made during the registration of the census. The fields FOLK.FODFORS, FOLK.FODORT, FOLK.FSCBKOD, FOLK:FODLAN and FOLK:FODLANKOD have all been given their values during the registration. When the NAPP codes are set all this information is used and compared to existing Swedish parish lists in order to produce the most accurate geographical coding possible. The geographical coding has been done by Johan Gidlöf at the Stockholm City Archives.

The geographical coding work will provide correct values for the variables **BPLCNTRY, NATNALTY** and **NAPPSTER.** The variable **BPLSE** contains a two digit county code that is stable over time. This is the same code used in the variable **COUNTYSE** belonging to the variable group "Geographic Variables (Household)". The variable **BPLPARSE** contains a 9-digit geographical code that is stable over time. This is the same code used in the variable **PARSE** belonging to the variable group "Geographic Variables (Household)". The variable

BPARSE contain the birth parish as transcribed and the variable **BRESSE** contain the birth place of residence as transcribed.

Ethnicity and Language Variables (Person)

The variable **ORIGIN** containing codes for ethnicity can be created for the Swedish 1880, 1890, 1900 and 1910 censuses. It is however one of the most problematic and complicated variables, because it has no direct correspondence in the Swedish FOLK-database. The ethnicity coding has been done by Per Axelsson and Isabelle Brännlund at Umeå University.

In the field list for the table FOLK (see chapter 3) the fields STAM (a field for "ethnic group"), UTKYRK, UTKYRKT and EJDOP have the remark "1900-field". These fields were added to the registration form for the Swedish 1900 census. In the earlier censuses (1860, 1870, 1880 and 1890) these fields didn't have their own headings. These fields were available in the registration software for the 1900 census but they were never used. The information found under these headings has instead been put in the field FOLK.OVRIGT. This field contains "additional information", that is information that doesn't fit in any other field.

The information about ethnicity has to be pieced together from several different fields in the FOLK-table. Ethnicity in the Swedish 1880, 1890, 1900 and 1910 censuses can be identified in at least six different ways:
1. Name or abbreviation for ethnic group in the field FOLK.OVRIGT. It can say "lapp" or just "l." for "Sami" and "finne" or "f." or "Finn".
2. Name or abbreviation for ethnic group can also (wrongly) be found in the field FOLK.NATIONAL for nationality.
3. Membership in an ethnic parish. The sami people were in some parts of Sweden recorded as members of special "sami parishes" called "lappförsamlingar" in Swedish.
4. Inhabitant in a place of residence with an ethnic name. Some villages were called "lappby" as a part of their name. The word "lappby" means "sami village"in English.
5. A person may have an ethnic occupation or title. A person described as "nomad" or "reindeer owner" is almost always a member of the Sami people.
6. Membership in a family unit where some member of the family unit has been given an ethnic coding according to the rules 1-5. There is some work to do regarding how family members should be coded. Mixed families must also be taken into consideration. In the northern county of Norrbotten the same little village may have Swedish, Finnish and Sami inhabitants.

The variable **RELIGION** can be created for the Swedish censuses. It is however one of the more complicated variables. The information has to be pieced together from several fields. Religion may be identified in four different ways:

1. Name or abbreviation for religion in the field FOLK.TROSB. The field contains information in text about religion, but it is not the only relevant field.
2. Information in the field FOLK.OVRIGT. As described earlier the information from the fields UTKYRK, UTKYRKT and EJDOP have been registered in the OVRIGT-field. This means that the field OVRIGT may contain information

about why a person has left the Swedish church and to what religion he has converted.

3. Membership in a "non Lutheran" parish. In some of the larger cities there are special Catholic parishes and Jewish parishes.
4. Membership in a family unit where some member of the family unit has been given a religion coding according to the rules 1-3. There is some work to do regarding how family members should be coded. Mixed families must also be taken into consideration.

In addition to the variable **RELIGION** the original plain text from the field FOLK.TROSB is saved in a variable **RELIGSE.**

Work Variables (Person)

The coding of occupations has been done by Carl Szabad at the Swedish National Archives in cooperation with Johan Gidlöf at the Stockholm City Archives.

There is a problem with how information about occupations have been registered in the Swedish Census-database FOLK. In the first registration DOS-software, used mainly for the 1890 census, all occupation information was transcribed directly from the source. When a Windows-software for registration was taken into use this changed. In the new software occupations could be chosen from two "drop-down-fields", one for "occupation name" and one for "occupation code". The chosen values could however be overwritten manually in the registration program and in the database there are many more unique "occupation string" values than there are values in the "drop-down-list". The shifting principles for how occupations have been registered are a problem, but we still believe that the occupation coding will be meaningful and useful to researchers.

The variable **OCCSTRNG** gets its value from the corresponding Swedish field FOLK.YRKE. The variable **OCCSE** contains the Swedish occupation codes from the field FOLK.YRKKOD. These codes were used by Statistics Sweden at the time of the 1880, 1890, 1900 and 1910 censuses and are used in the Swedish occupation statistics of the period.

The variable **OCCHISCO** contains the NAPP-HISCO code. The variable **OCRELATE** contains the NAPP-HISCO relationship code and the variable **OCSTATUS** contains the NAPP-HISCO status code. These codes have been set according to the NAPP-HISCO specification.

HISCO 2002 variables

In addition to the NAPP-HISCO variables there are also three variables that follow the HISCO 2002 standard. The variables **HISCOSE, HISCRELSE** and **HISCSTATSE** follows the guidelines in M.H.D. van Leeuwen, I. Maas and A. Miles. (2002) *HISCO: Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.

HISCOSE have been coded using the information in **OCCSTRNG** and **SEX**. This information has been matched against already established HISCO-codes

present at the Stockholm City Archives in Sweden (The Roteman Database HISCO-codes). The results of the matchings have been refined using manual coding and crosschecking against common Swedish codes published in the HISCO-manual mentioned above.

Users interested in variation in occupational responses below the level of detail provided in the codes (e.g; distinguishing "caretakers" from "janitors") should look at the **OCCSTRNG** variable. This is a Swedish language string variable.

HISCO is based on the International Standard Classification of Occupations from 1968, commonly known as ISCO-68.

Structure of the classification scheme
The classification scheme is hierarchical, in the sense that each digit in the 5 digit codes introduces a new level of detail. Codes sharing the same first 1, 2 or 3 digits are considered to be increasingly similar. For example, all people working in agriculture have the first digit 6. The first digit of a code indicates the "Major Group" a person's occupation is in.

The second digit indicates a "Minor Group" distinction. Continuing the previous example, people who have the first two digits "61" are farmers - who may specify what they are cultivating or tending - and farm managers. Thus, as well as sharing the characteristic of working in agriculture (6) they also share the characteristic of being owners or managers (61).

The first 3 digits denote the "Unit Group" of an occupation. At the third digit level, we introduce more detail. For example, the unit group "612" indicates "Specialized Farmers". Within this unit group, 4th and 5th digit distinctions known as "titles" or "headings" are made. For example, 61220 indicates "Field crop farmers," and "61230" indicates "Orchardists and fruit farmers."

Understanding headings
In general, if the last two digits of a code are "00" the heading is reserved for general titles. For example, in unit group "721" for "Metal smelter and furnace workers", the heading "72100" is reserved for responses such as "Metal smelter" and "Furnacemen."

If the last two digits of a code are "10" the heading is reserved for "not further specified" titles (often abbreviated "n.f.s." or "nfs" in syntax files), and codes ending in "20", "30" and higher multiples of 10 are reserved for responses with more detail on some aspect of the occupation. For example, 58210 is the code for "Policemen and detectives, employer unknown," whereas 58220 and 58230 are the codes for "Policemen and detectives, public service," and " Policemen and detectives, private service" respectively.

Headings ending in digits other than "0" (e.g; 2,3,4 . 9) are generally reserved for frequent responses specific to a particular country that probably belong with responses sharing the same first four digits of the heading. For example, 61115

(Husbandman or cottar) and 61117 (Female farmer) could be classified with other general farmers (61110), but occurred frequently enough in Norway and Canada respectively that we felt they should be given a separate code.

If the last two digits of a code are "90" the heading is reserved for "not elsewhere classified" responses. For example, 58290 is the code for "Other law enforcement officers."

The difference between "n.f.s." and "n.e.c." responses is that n.f.s responses are quite general, and may not offer much detail on the tasks and duties of the job. Conversely, n.e.c. responses are typically quite detailed in the information provided, but there are not enough similar responses to justify a separate heading.

Other HISCO 2002 work variables
Responses to the occupational questions also returned information that cannot be classified within an occupational classification scheme, such as relationship to workers, or indications of status. This information has been coded in the two variables **HISCRELSE** and **HISCSTATSE.**

HISCSTATSE preserves inconsistently available information in responses to occupational questions. The purpose of **HISCSTATSE** is to preserve information about the social status of an occupation. Correct usage of **HISCSTATSE** is critical to proper interpretation of the occupation codes.

HISCRELSE preserves inconsistently available information in responses to occupational questions. The purpose of **HISCRELSE** is to retain information about social and familial relationships around a person's occupation.

For example, some respondents indicated that they were a "laborer's wife." In this case the response received the occupational (**HISCOSE**) code for "laborer", and the **HISCRELSE** code for wife (11) to indicate that they were related as a wife to a worker.

Common occupations
Some very common occupations need special consideration because of the large number of records that are affected by the choice of HISCO codes. The following table contains a list of some of the most common occupations and their HISCO codes in the 2011 edition of the Swedish 1890 and 1900 censuses.

Table 12: HISCO codes for some common occupations

OCCSTRNG	HISCO 2002	RELATE 2002	STATUS 2002	HISCO NAPP	RELATE NAPP	STATUS NAPP
Dräng	62105	0	33	62110	0	34
Piga	54020	0	33	54020	0	34
Hemmansägare	61110	0	11	61110	0	14
Torpare	61115	0	12	61115	0	15
Inhyses etc	99999	0	13	99999	16	13
Arbetare	99900	0	32	99150	0	32

Source: The Swedish 1890 and 1900 censuses, Swedish National Archives

When the occupation coding was done for the first time in 2008 the occupations
"dräng" and "piga" were given the code 99998, indicating that more then one
interpretation was possible. In the Swedish census editions from 2011 and
onwards the coding has been done as shown in the table above. The occupation
"dräng" has been coded as 62105 "Farm worker, general" and the occupation
"piga" has been coded as 54020 "House servant". It could be argued that a woman
with the occupation "piga" in a rural area also functioned as a "farm worker", but
she usually performed some domestic work as well. The table also show the
coding for the very common occupations "hemmansägare" (farm owner) and
"torpare" (crofter).

The occupation strings "Inhyses", "Inhyseshjon" och "Fattighjon" (dependent
tenant) have all been coded with the HISCO code 99999. In 2008 the occupation
string "Inhyses" was coded with the HISCO-2002 code 99920 "Day labourer"
and with the status code 0. From 2011 and onwards the assumption has been that
people labelled as "inhyses" more often than not were poor people without an
occupation. The occupation "Arbetare" has been given the status code 32. In 2008
"arbetare" was coded with status code 0. This means that many more persons have
been given the HISCO status code 32 in the later editions than in 2008.

HISCO relation codes

The HISCO relation variable has been coded in more or less the same way in the
later census editions as it was in 2008. The most important difference is that the
NAPP-HISCO code 10 "Wife" has been used from 2011 and onwards. In HISCO-
2002 the relation code 11 corresponds to the NAPP-HISCO codes 10 and 11.

HISCO status codes

More occupations have been associated with status codes in the later editions
compared to when the coding was done for the first time in 2008. The main
difference is that very common occupations like "Dräng", "Piga", "Arbetare" and
"Inhyses" have been given status codes from 2011 and onwards. None of these
occupations had a status code in 2008. There are also a number of less frequent
occupations that have been given status codes in the later census editions.

Labor force

LABFORCE is a dichotomous variable identifying whether a person participated
in the labor force. For the Swedish 1880, 1890, 1900 and 1910 censuses
membership in the labor force has been defined on the basis of a person's age and
the values in the other HISCO variables. A person has been coded as a member of
the labor force if he has an occupation string in **OCCSTRNG** and is more than 15
years old and has no value in the **OCRELATE** variable and do not have the
values 40, 51 or 52 in the **OCSTATUS** variable. The codes for **LABFORCE** are
listed below:

Code:	Explanation:
1	No, not in the labor force
2	Yes, in the labor force
9	Unclassifiable (this code is not used in the Swedish censuses)

Multiple occupations

In many cases persons may have multiple occupations listed in the occupation string variable **OCCSTRNG**. The variable **OCCMULTISE** will contain a value that indicates if more than one occupation has been recorded for a person. The values will be as follows:

Code:	Explanation:
0	The person has no occupation or only one occupation.
1	The person has more than one recorded occupation.

Disability Variables (Person)

The Swedish variable **DISABSE** contains the information in the field FOLK.LYTE. The field contains text information about disabilities. The variable **DISABCO** contains an encoding of the disabilities and sicknesses in the field **DISABSE**. The value list is as follows:

Code:	Explanation:
20	Epilepsy.
21	Mental disease, Insanity
22	Mental retardation, Idiocy
24	Blind
25	Deaf
26	Deaf-mute (or mute)
98	Other
99	No information in source

The disabilities 20 to 26 in the code list are those that were explicitly asked for in the instructions for the census. Many more disabilities were recorded but they have all been coded as 98.

Many people have more than one disability. There are quite a few cases where a person has two or three disabilities. These extra disabilities are coded in the variables **DISABCO2** and **DISABCO3**. These variables share the same code list as the variable **DISABCO.** The variables **DISABCO, DISABCO2** and **DISABCO3** can be used to produce the integrated variables **BLIND, DEAF, IDIOTIC, INSANE** and **DISABIL.**

The NAPP variables for disability has changed somewhat since the disabilities for the Swedish 1890 and 1900 censuses was coded in 2011 but we have decided to keep the same variables and the same code values in the 1880 and 1910 censuses so that all Swedish censuses are compatible. The variable names and code values may be changed when the censuses are processed and published by the MPC.

Other Variables (Person)

The new variable **TITLESE** will contain the information from the field FOLK.TITEL. The variables **NAMELAST** and **NAMEFRST** contain the information from the corresponding fields FOLK.FORNAMN and FOLK.ENAMN in the FOLK database.

The variable **NOBAPTSE** contains a code stating if the person hasn't been baptized in the Swedish church. Many Baptists didn't baptize their small children in the church and this was recorded in the census. The value list is as follows:

Code: Explanation:
0   No information in the source
1   The person has the notation: Not baptized in the Swedish church.

The variable **CHEXITSE** contains a code for all persons that have left the Swedish Church. The value list will be as follows:

Code: Explanation:
0   No information in the source
1   The person has the notation: Has left the Swedish Church.

7   Notes on the Swedish censuses

The Swedish 1880, 1890, 1900 and 1910 censuses contain the same variables and the code book is the same for all four censuses. There are however a few differences between the censuses that has to be noted.

- There are many more unique phrases for family relations, disabilities and other variables in the 1890 census compared to the 1900 census. This means that even if some phrases occur both in 1890 and 1900 there were a substantial amount of new phrases that had to be encoded for the 1890 census.

- There is a difference in how households are enumerated in the 1890 census compared to the other censuses. In the 1880, 1900 and 1910 censuses households are indexed from 1 to N within each parish (there are however a few exceptions). In 1890 it is not unusual that households are indexed from 1 to N within each place of residence (**RESNAME**). The software used to produce the NAPP encoded files had to be modified to deal with this problem.

- In general there are more inconsistencies and faults in the 1890 census. The 1890 census was the first census registered and the registration practices for the 1890 census were more varying than for the 1900, 1880 and 1910 censuses which were digitized and registered later.