GRANT APPLICATION:

PROGRAMME FOR NWO LARGE INVESTMENTS

Life courses in context.

A collaboratory based on Dutch population registers and censuses (19th and 20th century).

Royal Netherlands Academy of Arts and Sciences (KNAW)
International Institute of Social Science History (IISH)
Netherlands Institute for Scientific Information Services (NIWI)

21th August 2001

1. SUBMISSION, TITLE OF APPLICATION, APPLICANTS, SUPERVISORY BOARD, MANAGEMENT AND EXECUTIVE SUMMARY

1.1 Submission

The application is being submitted to the Governing Board and relates to the Humanities and the Social Science Councils.

1.2 Title of application

Life courses in context. A collaboratory based on Dutch population registers and censuses (19th and 20th century).

1.3 Applicants

Royal Netherlands Academy of Arts and Sciences (KNAW)

on behalf of

International Institute for Social Science History (IISH)
Netherlands Institute for Scientific Information Services (NIWI)

in collaboration with

Foundation Historical Sample of the Population of the Netherlands (HSN)
Foundation Historical Database of Dutch Municipalities (HDNG)
Statistics Netherlands (CBS; Central Bureau of Statistics)

1.4 Academic Supervisory Board

- Dr Th. Engelen, historian, University of Nijmegen, Historical Database of Dutch Municipalities
- Prof. H.B.G. Ganzeboom, sociologist (University of Utrecht), Chairman of the Historical Sample of the Population of the Netherlands (HSN)
- Prof. A.F. Heerma van Voss, historian (Associate Director of Research, International Institute of Social History, University of Utrecht), Secretary of the HSN steering group
- Dr H. Knippenberg, historical geographer (University of Amsterdam), Vice-Chairman and Treasurer of the HSN Foundation
- Dr J. Kok, historian (postdoctoral researcher, International Institute of Social History), Board member of the HSN Foundation
- Dr M.H.D. van Leeuwen, historian and sociologist (postdoctoral researcher, Netherlands Economic History Archive), Board member of the HSN Foundation
- Prof. J.M.W.G. Lucassen, historian (International Institute of Social History (IISH),

Free University, Amsterdam, Board member of the HSN Foundation
- Dr I. Maas, sociologist (postdoctoral researcher, University of Utrecht), Board member of the HSN Foundation
- Dr J. van Maarseveen, Statistics Netherlands, Voorburg, Historical Database of Dutch Municipalities
- Prof. F.W.A. van Poppel, historical demographer (Netherlands Interdisciplinary Demographic Institute, University of Nijmegen), Secretary of the HSN Foundation
- Dr H. Wals, historian (Deputy Director, International Institute of Social History), Member of the HSN steering group

See Appendix A1 for a list of relevant publications by members of the board.

1.5 Executive management

- Dr P.K. Doorn, Head of Historical Information Services, Netherlands Institute for Scientific Information Services (NIWI)
- Dr C.A. Mandemakers, Head of the Historical Sample of the Population of the Netherlands (HSN)
  Senior researcher, International Institute of Social History (IISH)

See Appendix A2 for a list of relevant publications by members of the management.

1.6 Address of correspondence

Dr. C.A. Mandemakers
International Institute of Social History
Cruquiusweg 31
1019 AT Amsterdam
The Netherlands
tel. +31 20 66 858 66
fax: +31 20 66 541 81
email: KMA@iisg.nl

1.7 Executive summary

The programme's objective is to develop a database with about 40.000 individual life courses of people born in the period of 1863-1922. This database with microdata will be supplemented with core data on the level of the municipalities. This will be done by digitalizing the results of the ten-yearly censuses as they were taken between 1859 and 1947.

The programme will build on work already carried out by two KNAW institutes – the International Institute of Social History (IISH) for the Historical Sample of the Population of the Netherlands (HSN) and the Netherlands Institute for Scientific Information Services (NIWI) for the digitised census data– in collaboration with scholars active in a broad group of research institutes: N.W. Posthumus Institute (Netherlands Research Institute and Graduate School on Economic and Social History), Ceres (Research School for Resource Studies for Development), The Interuniversity Center for Social Science Theory and Methodology (ICS), The Netherlands Interdisciplinary Demographic Institute (NIDI) and Statistics Netherlands (CBS).

Contrary to some other Western countries census data on individuals for the period before 1960 do not exist anymore for the Netherlands. To bring research on comparable international standards it is necessary to reconstruct these data. But, the construction of this database will not only eliminate the disadvantage of lacking data on the micro level, faced by the Netherlands compared with other countries. The HSN database, drawn from the Dutch population registers, *will actually put the Netherlands in a unique position in the international research world*. Because the population registers have recorded details of every change of residence for every individual since 1850, investigators can have access to research populations that are not limited to persons who stayed put in just one municipality. Dutch researchers who know the Dutch situation and sources will have a priviliged access to unique materials (which will off course also be available to non-Dutch scholars).

The database of the Historical Sample of the Population of the Netherlands (HSN) will cover the entire country and contain micro-level data on the life courses of over 40,000 individuals born between 1863 and 1922. These life courses are to include data on each successive family situation in which the individuals lived, all the addresses where they lived, as well as data on the religion and occupational title of each subject and of every person with whom they co-resided (and, for married subjects, data on the occupational title and place of residence of family members of the subject's spouse). In the social and historical sciences, the life-course perspective is increasingly important in explaining demographic and socio-structural processes. The HSN database will cover the entire period from 1863 to 2000 and can be regarded as a chronological expansion of the retrospective databases currently in use by the social sciences which database consists of data form persons only born in the period after 1900 (Liefbroer & Dykstra 2000) and which have very scarce information about the parental situation for the earlier cohorts.

Individual life courses have to be analysed in a rapidly changing environment of a industrializing and modernizing Dutch society. The Dutch national censuses form a fundamental source of information for conditions on the level of the municipality. In addition to the population size, population censuses contain information on the structural characteristics of the population, such as age, gender, marital status, religion, household status, occupational activity, and nationality. In some years the censuses were combined with an occupational census and a housing census. The present grant application includes the digitisation of the population and occupational censuses of the Netherlands for the full period 1859 to 1947. The data will be linked and integrated with those from the Historical Database of Dutch Municipalities and will build on those censuses already compiled by the

Netherlands Institute for Scientific Information Services (NIWI) and Statistics Netherlands (1795-1859 and 1899).

The expansion of the existing databases of HSN and NIWI into complete systems will be of major significance in the long term. In the short term it will provide a strong stimulus to academic research into social, demographic and economic developments over the past two centuries. Based on these databases, a research programme, here attached as appendix D, has been drawn up involving the principal research groups using quantitative material on the Netherlands in the nineteenth and twentieth centuries. This programme will result in a series of studies in historical demography, social and economic history, human geography, sociology and epidemiology. The work will be performed in stages so that the data will become available for research before the project is completed.

Another aspect of methodological innovation envisaged by this programme lies in the scope for connecting data from the micro and meso/macro levels. The population censuses and other municipal data offer a context for the individual-level and family-level data. The combination of the different sources will create new opportunities for multi-level or cross-level analysis.

Digitisation of the population registers and the population censuses is one of the first priorities stated in the NWO trend report *Een digitale bibliotheek voor de geesteswetenschappen* [*A digital library for the humanities*] (Viskil 1999). The results of the propesed project will be presented and communicated in the form of a collaboratory for the humanities. A collaboratory is an open meta-laboratory that spans multiple geographical areas with collaborators interacting via electronic means - "working together apart." The impact of these databases will however not be confined to the research world. The databases will be accessible on the Internet to the general public. However, for privacy reasons files or data relating to individuals still living will only be accessible using a technique known as *remote execution.*

2. ORGANISATIONAL STRUCTURE OF THE PROGRAMME

Since 1991 the Historical Sample of the Population of the Netherlands (HSN) has been working at the International Institute of Social History to construct a database containing micro-level data on the Dutch population. And since 1997 the Netherlands Institute for Scientific Information Services (NIWI), in collaboration with Statistics Netherlands (and more recently with the Historical Databank of Dutch Municipalities), has been digitising the census data published in countless volumes of official statistics. The work is thus being carried out under the auspices of two KNAW institutes. Through these projects, the institutes have built up substantial relevant experience in this field in recent years, and both institutes have been able to interest or recruit researchers from various universities to take part in these projects.

Work on the proposed programme will commence on 1 October 2002 and be completed

by 30 September 2007. The application covers the costs of acquiring, entering, managing and making available the data, as well as the costs of co-ordinating the work during these five years. Overall co-ordination of the programme will take place in regular consultations between the two proposed executive managers: Dr P.K. Doorn for the population censuses and Dr C.A. Mandemakers for the life courses. To monitor the progress of the project and ensure its academic quality, an academic supervisory board will be appointed, comprising individuals who have been, or still are, closely involved with, and responsible for, the various databases (see section 1.4). The board will meet at least three times per year.

For a selection of all research which can and will be undertaken, we refer to the research programme which paper accompanies this application (appendix D). Both IISG and NIWI will devote 1.0 fte for coordinating research on the basis of the databases of life courses and census.

The work on the databases will be organized into different stages. After two years the dataset will be complete for the persons born in the provinces of Friesland, Utrecht, Zeeland and the city of Rotterdam and at that moment also the census data will be ready for use. After three years the dataset will be complete for all persons born between 1885 and 1922. Besides this schedule, further specific releases of data will be responsive to the needs of researchers.

The contents of this application is as follows: Chapters 3 and 4 present an overview of the current situation regarding the two core databases. They provide more detail and arguments for the proposed expansion. Chapter 5 outlines the budget and the timing of the workflow. The detailed budget of the work is included in Appendix B1 and B2. Chapter 6 discusses the way results will be presented and communicated ('collaboratory for the humanities') and the academic relevance of the programme (including its national importance, and nationwide accessibility), chapter 7 highlights the international significance of these databases, and chapter 8 discusses the future of the proposed databases within and outside the applicant institutes (organisational strategy, and operational investment).

3. LIFE COURSES (HSN)

3.1 The HSN

The proposed life-history database will be based on the existing HSN database. The HSN aims to reconstruct life courses as completely as possible for a representative segment of the nineteenth and twentieth century population. The sample required for this purpose (N=77,000) has been derived from the birth registers for the period 1812-1922. Earlier grants from the Programme for NWO Medium-Sized Investments have already enabled this base sample to be input. A large number of death certificates and all the personal record cards of the subjects in the sample have also been added to the database. Work is currently underway to input marriage certificates and the information given in the population

register when the sample subjects were born. Much diverse research has been and is still being carried out using the database. Research based on the HSN has so far resulted in a large number of publications in the Netherlands and abroad including two PhD theses.

The dynamic population registration system introduced in the Netherlands in 1850 has resulted in an archive that is unique in international terms. It is unique because it recorded details of where inmigrants came from and where outmigrants went to, thus enabling the inclusion of the complete life courses of *migrants* in the dataset. The present application thus involves the systematic retrieval of the data on life courses from the population registers. Although the population registers were already in use in 1850, we will input life-course data only for subjects born in 1863 or later. There are two reasons for this: first, the registers did not function properly everywhere in the Netherlands during the first few years of their existence. Second, new regulations concerning population registration were introduced in the course of 1862. As a result, the design of the system was modified, and every household was then re-registered (Knotter & Meijer 1995).

Once the population registers have been input, the HSN database will contain the following information for the 40,000 subjects born after 1862: the occupational history of the subject, the marital history and family history of the subject (including age at marriage, religion, number of children), the parents of the subject (occupation, age at death, and cause of death (if available), place of birth, and an indication of illiteracy (inferred from the absence of signatures)), social network (the same data for witnesses at marriages), household dynamics (including care arrangements), and the migration history of the subject. These features will make the dataset a fundamental source for scholars investigating historical issues in demography, sociology, epidemiology, social economics and human geography. The nationwide coverage of the dataset ensures that regional variations can be identified, whereas current research involving such topics is usually local in nature and necessarily excludes migrants.

3.2 The database

For the HSN database, data for each individual are systematically collected from the records kept in the public archives. Primarily, these are birth certificates, death certificates, and personal record cards. The birth certificates include information on the person born, as well as the names, addresses, ages and occupations of the parents. The death certificates include the most recent place of residence and most recent occupation of the deceased, and information on his/her spouse(s); certificates of deceased children provide a second indication of the occupational title of the father (the person reporting the death) as well as a double-check on illiteracy. The personal record cards for all subjects who died between 1 January 1940 and 1 October 1994 have now been input. The cards include data on occupation (from 1940), cause of death (up to 1953), a full migration history (all addresses), family composition, and religion. On the basis of these data, it is now already possible to research topics such as childhood mortality and migration patterns for the whole of the Netherlands.

The above data are now to be supplemented by information from marriage certificates. These give details of the occupational titles, literacy (signature), and place of residence of the bride and bridegroom, their parents and the witnesses (usually friends or family of the couple). These certificates will enable scholars to research topics such as social and geographical mobility, marital mobility and literacy. A substantial proportion of the marriage certificates in the provinces of Utrecht, Friesland, Limburg, Gelderland, Groningen and Zeeland, have already been entered. Much useful "pilot" experience has been gained through the data input. Furthermore, various studies based on the certificates have already been published.

At a later stage, information from the population registers, land registers and tax records will be input. These sources are extremely rich, providing information on the family structure, pattern of migration, further occupational history, and the income and wealth of the subject (and sometimes of his or her relatives). From 1850 the Dutch population registers were maintained as dynamic records. By this we mean that the registers did not merely record a situation at a particular moment in time (a snapshot), but that all changes in a subject's address, family size and migration are noted, creating a longitudinal record. From 1870 onwards the records are actually fairly accurate. Many subjects born after 1870 can also be found in the personal record cards archive at the Central Bureau for Genealogy, so that their migration history can be traced in the reverse direction, thus minimising the risk of "losing" a subject.

The subjects were selected by taking a simple random sample from the birth registers for 1812-1922. The aim was to secure a sample size of 77,000. This is just over half of one per cent of the total number of births, assuming around 14.5 million people were born in the Netherlands in this period. A sample of 77,000 is sufficient for drawing statistically reliable conclusions on subpopulations of two per cent or more of the population born in the Netherlands during the period.

In addition to being an important source for research and a control database that can be used in interpreting findings on specific groups, the HSN database also acts as a foundation for the collection of new data. In practice, this is achieved by maintaining a data structure that can be used by individual researchers, and by consistently using the database as a starting point in subsequent research, both by expanding the number of subjects included (oversampling) and enriching the database by introducing supplementary data for specific groups of subjects. For researchers, it cuts both ways. Not only can they use the material already input, they also have access to the software and expertise developed by the HSN. This expertise can be seen as an important by-product of the data-entry work carried out over the past ten years. In return for the use of the software and the data already recorded, the HSN requires researchers to add to the dataset any new data they collect in the course of their research, thus ultimately making it available to other researchers too.

Public access to the data is subject to a dedicated set of privacy regulations (Dutch Data Protection Authority, number O-0030426; Law for the Protection of Personal Data: Wet Bescherming Persoonsgegevens, 6 juli 2000, Stb. 302). In accordance with the Personal Data Protection Act, the premise embodied in these regulations is that public access to

data for research purposes is governed by the same arrangements as those of the archive from which the data were derived. This might mean that some data can only be made available in an anonymised form. The work on the database is carried out at the International Institute of Social History; ownership of the data rests with the HSN Foundation.

3.3 Current size and expansion

The sample contains almost 77,000 persons. So far (at date 1st June 2001) the following have been entered: birth certificates for the entire group, 22,000 death certificates, 16,000 personal record cards (available only for persons alive on 1 January 1940), 9000 marriage certificates, and details of 4000 initial registrations in the population register. The aim is to add a further 13,000 marriage certificates and 12,000 population-register registrations by 31 December 2002.

Through the collaborative projects with other researchers, the HSN database has been further enriched with around 20,000 birth certificates, 4000 marriage certificates, 4000 personal record cards and information from the population registers for 5000 individuals.

In general, one can assume that the research community will benefit from having life courses that are as complete as possible. Including data from the population register and complementing these with data from marriage certificates will give us the following information (for a comprehensive overview, see appendix C4):

- composition of the family into which the subject is born, and the changes in that family before the subject ultimately leaves home
  illiteracy of the subject's father (evidenced by the absence of signature on birth certificate - or on death certificate of the subject)
- migration history, including information on boarding houses, etc.
- occupational title, marital status and religion of all the relatives with whom the subject co-resides
- occupational title of the parents, parents-in-law, four witnesses, subject and partner of the subject (in the event of a marriage certificate)
- literacy in the subject's social environment (signature of parents, parents-in-law, four witnesses, subject and partner of subject (in the event of a marriage certificate))
- composition of the subject's own nuclear family and the changes in that family prior to the subject's death
- relief or care arrangements for the subject in old age.

Ensuring that life courses are as complete as possible means including in the dataset the marriage certificates and death certificates not yet added. Besides serving as a check on population register data, marriage certificates also provide evidence of social mobility and literacy, since they include data on place of residence, occupation, age, literacy (evidenced by a signature) of bride, bridegroom, their parents and four witnesses (usually relatives or close friends of the bride and bridegroom). Much international *comparative* research uses marriage certificates, because many countries that lack a population register do have

marriage certificates. Several articles in international journals have already reported research based on the HSN material (e.g. Van Leeuwen & Maas 1997, Delger & Kok 1998).

This dataset is a fundamental source for scholars engaged in historical research in demography, sociology, epidemiology, social economics and human geography. The overview of joint matching projects included in table 4 in appendix B1 summarises the research already carried out with the HSN database. For details regarding the HSN in a long-term perspective, see chapter 8.3.

4. CENSUS DATA

4.1 The population census

National population censuses are one of the fundamental sources of information on conditions in a country. In addition to the population size, the population census generally contains information on the structural characteristics of a population, such as age, gender, marital status, religion, household status, occupational activity and nationality. In some years the Dutch censuses were combined with an occupational census and a housing census.

The first general Dutch population census was held in 1795 under the Batavian Republic. From 1829 onward, censuses were held every ten years. The 1940 census was postponed until 1947 because of the war. No population census has been held in the Netherlands since 1971 because of growing privacy consciousness (and refusal to take part) among the general public.

Only a limited number of original copies of the 200 or so published volumes of the 1795-1971 Dutch population censuses have survived. These censuses have always played an important part in historical and social-science research. Many of the published census volumes are now in poor condition. Digitisation can help preserve this material while also increasing access to it. The *c*. 42,500 pages contained in these volumes have now been scanned and are available digitally on CD-ROM. About 35,000 pages relate to the period 1859-1947. About 10,000 pages of the 1899 census have been published electronically and about 5,000 more pages have been converted to digital form but are not yet available. About 20,000 pages of tables still need to be converted in order to make all published population censuses until 1947.

The digital census databases will be important resources for historical and social-science research. National historical census projects have been, or are being, carried out in a number of countries, including the US, the UK, Ireland, France, Norway, Denmark, Germany, Russia and Austria. Some of these projects are based on the original source material, which means the databases can be constructed at the level of the individual.

4.2 Current situation regarding digitisation

Between 1997 and 1999 the CBS and the Netherlands Institute for Scientific Information Services (NIWI) collaborated on a project entitled "Digitising the 1795-1971 Dutch Population Censuses". This project has been successfully completed and has resulted in two sets of CD-ROMs, a joint website, an academic symposium, and a book containing the results of the symposium. The press has been extremely enthusiastic about the project. The 750 sets of CD-ROMs produced are almost sold out, and the website is used intensively. Researchers have also been enthusiastic about the results, and particularly about the unabridged electronic publication of the 1899 census. The programme now being proposed will fully digitise all the censuses between 1859 and 1947. So far, every published census volume (*c*. 200 volumes containing around 42,500 pages) has been made accessible, but only in the form of digital images. Only the 1899 census (including the occupational and housing censuses, 26 volumes, *c*. 10,000 pages) now has a comprehensive database that can be searched and analysed (accessible via CBS StatLine).

4.3 Full digitisation

The purpose of this part of the programme is to construct and make accessible a comprehensive database containing the 1859-1947 national population censuses (including occupational and housing censuses). The population censuses, which form the basis for so many other statistics, will be a key research resource for social and economic historians, demographers, social scientists and epidemiologists. Secondary target groups are amateur historians, local authorities, the media and education. The project will comprise the following elements:

- Making accessible the figures from the 1859 and 1930 census publications that have already been input, and publishing them digitally on CD-ROM and the Internet and publishing them digitally on CD-ROM and the Internet; the software for retrieval and access to the data will be StatLine, a package developed by the CBS (Statistics Netherlands).
- Content conversion (by manual data entry and/or optical character recognition) of the figures from the 1869, 1879, 1889, 1909, 1920 and 1947 censuses (ca. 20,000 pages) and publishing them in digital form. The material needs to be checked and, where necessary, corrected, documented and converted to the retrieval StatLine. Wherever possible, the data will be standardised or uniformalised to enable comparisons over time.
- A historical database of co-ordinates is available, enabling data at the municipal level to be represented visually for any desired point in time during the past two centuries. The project will also provide additional information for the Historical Database of Dutch Municipalities (HDNG), which is currently being developed and in which the universities of Nijmegen and Amsterdam, the Netherlands Interdisciplinary Demographic Institute and the Historical Sample of the Population of the Netherlands collaborate. A number of existing historical-statistical and

demographic databases (the historical-ecological database at the University of Amsterdam, the "Hofstee" database at the NIDI, and databases at the University of Nijmegen) will be integrated with the HDNG.

5. BUDGET AND TIMING WORKFLOW

5.1 Budget

The following table summarises the budget for the entire five-year period. More detailed costings are included as Appendix B1 and B2. The budget does not include the costs of office space and administrative support. These will be borne by the KNAW and its institutes (the IISH and the NIWI). The amounts stated in the budget for these institutes relate to contributions to the personnel costs of the project. "Matching" refers to contributions to the programmes from other parties.

BUDGET	EUR (*1.000)
Life-courses	3740
Census digitisation	835
TOTAL COSTS	4575
Contribution from IISH	200
Contribution from NIWI	160
Contribution from third parties (matching)	530
TOTAL CONTRIBUTIONS	890
NWO INVESTMENT GRANT	3685

5.2 Time schedule workflow

Total workflow will be divided on the basis of the following lines

9) Researchers need as soon as possible complete and coherent data, this means that we will concentrate first on some regions and one specific period:
   - REGION Alle research persons (RP's) of which we have already collected some data: the provinces of Zeeland, Utrecht and Friesland and the city of Rotterdam ('core regions' vs. other regions). This part of the database will be ready for use after the second year.
   - PERIOD All RP's who have been born in the period between 1885 and 1922. Because life course data from this specific period are relatively easy to gather. This part of the database will be finished at the end of the third year.

2) Migration of Research Persons to other regions is very important (about 50 percent of all grown ups) but also costly in (waiting) time. This means that we have to reserve quite a long period for sorting out problems and waiting on results from work done by municipal archives. So from a logistic point of view the project can be distinguished in two phases: A) a relatively short period of processing all data in the municipality of birth and B) a relatively long period for handling all migration steps. This means that most of the work will be done in the first three years of the project.
3) The census database will be ready for use at the end of the second year of the programme when the first data of life courses will be released (phase C in the table below).

This results in the following time table:

Year	LIFE COURSE DATA				CENSUS DATA	Spending (* million EURO)
	REGION		PERIOD			
	Core	Other	1885-1922	1863-1884		
1	A/B	A	A		C	1.1
2	B	A	A / B	A	C	1.5
3		A /B	B	A / B		1
4		B		B		0.6
5		B		B		0.4
Total						4.6

Explanation

CORE REGION Zeeland, Utrecht, Friesland, Rotterdam
OTHER REGION Other provinces and towns
A Data processing of data from municipalities of birth

B	Migration steps
C	Work on censuses

6. NATIONAL SIGNIFICANCE AND NATIONWIDE ACCESSIBILITY

6.1 A collaboratory for the humanities

The presentation of the results and the communication with the academic world will take the form of a collaboratory for the humanities.

A collaboratory is an open meta-laboratory that spans multiple geographical areas with collaborators interacting via electronic means - "working together apart." Collaboratories are designed to enable close ties between scientists in a given research area, to promote collaborations involving scientists in diverse areas, to accelerate the development and dissemination of basic knowledge, and to minimize the time-lag between discovery and application. The collaboratory concept of an enabling collaborative environment has been applied to diverse areas including experimental and computational research in the biological and physical sciences, clinical medicine and medical information systems, computer and information sciences, mathematics and law.

The databases to be developed will form the core around which historical and social scientific research on the Netherlands in the past two centuries will take place. The databases will be open systems (as far as privacy matters and other data security considerations permit). Every researcher should be able to perform operations such as establishing their own links within the dataset, adding their own classifications, linking the data to other Dutch and foreign datasets, and to expand the existing database with their own research data.

6.2 Academic importance

The database to be constructed will be of great benefit to researchers in a wide range of disciplines. This is already evident from the fact that so many researchers are currently working with this type of data, or have done so in the past. However, up to now their methods have been highly time consuming and inefficient. Though population censuses and occupational censuses were recently made available in image form, to supplement the original volumes held by academic libraries, data still has to be searched and copied manually. A simple example: it would currently take at least a day to tabulate the number of medical doctors per province for each census year; after digitisation it will take just five minutes to select the data and less than one minute to tabulate the results.

Similar considerations apply to the life-history database. Over the past 20 years researchers from many disciplines have created databases on individuals from the nineteenth and early twentieth centuries. Almost all such databases contain only the data collected for one particular study; little has been done with them subsequently. Since the databases were created for specific studies it is often not worthwhile completing them, including, say, the life courses of subjects migrating to other municipalities. This was indeed one of the principal reasons for establishing the HSN in 1988. The construction of

one large core database of life courses will complete the work already started and will enable research to be done that was previously infeasible or severely restricted in scale and scope.

Since the databases are expected to be vitally important and heavily used, digitisation of the population registers and censuses is one of the first priorities stated in *Een digitale bibliotheek voor de geesteswetenschappen* [*A digital library for the humanities*], a policy memorandum published by NWO's Humanities Council in 1999 (Viskil 1999). The historians interviewed for this memorandum favoured the creation of nationwide databases, with the digitising of the population registers as a key priority.

Beyond these practical reasons, there are also more fundamental considerations that necessitate the completion of both databases. For many historical problems the microlevel of the individual and his/her family is the most appropriate level of investigation, but because such historical "survey data" are lacking, researchers are often obliged to seek answers at higher levels. A life-history database is admirably suited to answering many types of questions at the desired level. Both the census database and the HSN database will also enable us to situate highly specific or regional studies in their appropriate contexts and to generalise findings to larger groups. The information in the database will also permit more focused research questions, such as those in regional in-depth studies.

The facilities created by the programme will constitute integral parts of the national and international research infrastructure. This type of database transcends individual projects, providing interdisciplinary services for a wide variety of separate subsequent projects. Both databases will be suited for research questions in the historical sciences, demography, geography, sociology, economics, and the medical sciences. The research programme appended to the present application gives only a hint of the future research that will be possible with these databases.

The list of joint projects that accompanies the budget (table 4, Appendix B1) shows that over the past 10 years the HSN has also been a source of data and inspiration for research in widely varied disciplines, as reflected in part by the list of research schools (Posthumus Institute, ICS) and other institutions (NIDI, IISH, University of Nijmegen, etc.) with which the HSN co-operates. In this respect the HSN is a model of multidisciplinary collaboration.

The HSN database furthermore provides an infrastructure for the introduction of new data, such as that from tax registers, notarial and cadastral records, church archives, militia archives and military registers. Scholars carrying out research on this type of data can use the sample of individuals already selected by the HSN and the life-history information already available on these individuals for their own archive work, or alternatively use HSN data as a control database.

6.3 Social significance

Although the importance of both databases is primarily academic, they will also have a wider social impact. The databases will be publicly accessible, and will prove highly useful in fields such as university and secondary-school teaching. Dutch historical datasets will now be available for use in university instruction, for statistical exercises or as a reference framework. Secondary-school pupils could be given project assignments to compare local data with data of a more national character. In the case of life courses, schools could undertake research in local archives as a semi-permanent project, comparing the data collected with the HSN dataset (as long as this does not contravene the HSN's privacy regulations; see the following section).

6.4 Nationwide accessibility

Both databases will be made available through the channels normally used by the NIWI and the HSN. The NIWI will make a particular effort to inform potential users about the HSN database. The NIWI includes the Netherlands Historical Data Archive (NHDA) and the data archive for the social sciences (the Steinmetz Archive), both dedicated to preserving and disseminating large datasets for academic users. The Steinmetz Archive has a more social-science-oriented group of users, and has considerable experience in international collaborative ventures involving data archives (CESSDA, IFDO, IASSIST). The NHDA serves as an important gateway for Dutch and foreign historians.

In addition to the traditional methods of data dissemination – Internet sites (www.niwi.knaw.nl, www.iisg.nl/~hsn, www.volkstelling.nl, www.cbs.nl), CD-ROMs, tapes and disks – the information will also be accessible by remote access or remote execution. Remote access allows researchers to compile their own databases for downloading; remote execution is applied where privacy-sensitive twentieth-century microlevel data are involved. This technique makes it possible to conduct analyses without the danger that the anonymity of individuals is compromised.

The management and dissemination of the HSN data require further explanation. The HSN Foundation administers the HSN data, with the explicit task of making these available to researchers in the Netherlands and abroad. The confidentiality of the data is safeguarded by a dedicated set of privacy regulations (see Appendix C3). The basic premise of these regulations is that public access to data for research purposes is governed by the same arrangements as those of the archive from which the data were derived. In practice, this means that all nineteenth-century data can be provided free of restrictions. Data for the twentieth century will be anonymised. In cases where researchers require data on named individuals (this is necessary if they want to expand the dataset), a number of additional safeguards are built in to ensure confidentiality. The HSN Foundation also actively promotes access to the data by publishing information materials, organising conferences, and initiating research projects that utilise HSN data.

7. INTERNATIONAL SIGNIFICANCE

7.1 Sources unique to the Netherlands

Few countries in the world have a system of population registers like the Dutch one; even fewer had such a system in the past. The introduction of the population registers in the Netherlands in 1850, which enabled a continuous system of record keeping, makes it theoretically possible to trace every individual from birth to death throughout the country. Only Belgium, small parts of Italy, Norway, Sweden and Japan have historically comparable systems. Other countries have the original population census data at best; these are not continuous registrations but merely decennial snapshots.
The Netherlands also has registers of births, deaths and marriages. Although these are by no means unique to the Netherlands, they are notable for their ease of access and excellent potential for identifying individuals (and for linking to population register data).
For these reasons, the HSN data are of great interest to foreign researchers. The data collection methodology is also unique in several other respects. This can best be understood by first considering some foreign counterparts.

7.2 Foreign counterparts

Similar projects have existed outside the Netherlands for some time, particularly in Canada, Sweden and France. In Canada, two major projects are underway. One of them aims to reconstruct the *entire* Quebec population for the period 1608 to 1765. The second, at the University of Quebec at Chicoutimi, involves inputting into a database *every* marriage certificate issued between 1842 and 1971. So far, almost 1,5 million certificates have been entered.
Sweden has two large databases with life courses for a limited number of regions. The oldest is the Demographic Database based in Umeå, which focuses on northern Sweden. It now contains the life courses of 800.000 individuals, mostly for the nineteenth century. In addition, work is underway to construct a national database containing parish-level data, largely drawn from Sweden's population censuses. This makes the Swedish data collection in two respects an excellent model for the programme proposed in this application. A second major Swedish project, the Stockholm Historical Data Base, involves the input of all population registration data. This project falls under the auspices of the city archives and will eventually contain the entire Stockholm population register data for the period 1878 to 1926. Around 30% of the estimated 8 million entries have been processed so far.

Databases with raw data from nineteenth- and twentieth-century population censuses are of more recent origin. These censuses are becoming available digitally on a large scale, particularly in the US, Canada, the UK and Norway. In the US, the University of Minnesota maintains the IPUMS (Integrated Public Use Microdata Series) website. This gives access to microdata from a sample averaging 1% drawn from the US population censuses of 1850 to 1950. The data can be directly downloaded. Canada has the Canadian Families Project, whose objective is similar. Norway even intends to input all its nineteenth-century population censuses. In the UK, the 1880 census is currently being published electronically By the History Data Service at Essex.

The International Microdata Access Group (IMAG) has been set up to foster international collaboration between institutions and agencies inputting this census material. Other organisations, focusing on complete life courses, have also joined; they include the HSN and Umeå's Demographic Database, for a recent listing see the *Handbook of International Historical Microdata for Population Research* (Kelly Hall *et all*, 2000).

7.3 Uniqueness of the HSN

The HSN database differs from its foreign counterparts in combining the following three elements: a) the entire life course is documented, b) the research covers the whole country, and c) data are collected only for a sample from the population. The use of birth certificates as a sampling frame ensures it is reliable and representative.

The HSN is based on a research concept that treats the Netherlands as a single geographical area, despite the geographical diffuseness and local variations in the quality and accessibility of the research material. In some countries and regions one needs to input the entire archive to ensure that the data for at least some of the subjects are complete. This is not the case in the Netherlands, since there are many ways to recover "lost" subjects, for example by consulting the personal record cards archive at the Central Bureau for Genealogy and the indexes of births, deaths and marriages that are becoming increasingly available. In this respect the Dutch situation is unique, and the data are therefore of prime value for research at the international level. There is currently extensive research interest, nationally and internationally, in the migration of individuals, but that is an area where adequate datasets are notably lacking. This problem is also evident in research for which migration is not a direct concern but where it influences the results, as in studies on social stratification or health trends.

Despite its nationwide character, the database will be compact in size, since it is based on a random sample derived from birth registers. This is sufficient for most research topics. Only research that needs to make extremely detailed breakdowns would require an entire population to be included. The relative compactness of the sample will give the HSN an advantage over its foreign counterparts in terms of the resources and time required.

8. ORGANISATIONAL STRATEGY

8.1 General

Both institutes operate under the wider auspices of the Royal Netherlands Academy of Arts and Sciences (KNAW). The KNAW (like the NWO and SURF/IWI) regards the development of digital core databases for the humanities as a high priority for the coming years. The NIWI has been assigned a national co-ordinating role in this effort (see *Viskil, 1999* [*A digital library for the humanities*]).
The NIWI has a sound reputation when it comes to digitising data for the humanities and the heritage sector. Researchers in the humanities are the primary target group. In recent years the NIWI has built up specialist expertise in the digitising and digital archiving of historical sources and related heritage material in the humanities, and in making this accessible to users.
In its most recent policy documents, the IISH, too, has put special emphasis on digitisation. Both in the Netherlands and abroad, the IISH is regarded as a leader in the digitisation of archive material. The IISH is also active in research employing large historical databases and in providing the electronic infrastructure for this research, including portals, digital access to libraries, archives and collections, images, databases of research material, and source publications. As part of its commitment to a digital strategy, the IISH has provided a home to the HSN since 1991.

8.2 The NIWI

The NIWI's strategy is to strengthen the information infrastructure for researchers in history, Dutch language and literature, and the social sciences. In doing so it seeks to collaborate with research communities comprising foremost specialists in a particular field. Such communities might be national or international, and they might include both scholars and other interested parties (policymakers, cultural organisations, the media). Working with these communities, the NIWI endeavours to establish and maintain a comprehensive information infrastructure that reflects the needs and demands of researchers. The resulting expertise and infrastructure form the basis for new projects and core activities within the NIWI. In most academic disciplines such specific information infrastructures are now very much in the ascendant, bearing names such as "knowledge domain" and "information portal". Although such concepts are not yet clearly defined, in every case the aim is to create a single Internet portal for all the information and communication facilities relevant to the user group: databases, literature, discussion groups, software and other tools, news, employment opportunities, etc.

In many fields, research databases are becoming increasingly important as centres of such information infrastructures. Examples of current interest from the physical and life sciences include the Human Genome Project, the Human Brain Project, and meteorological databases. Similar developments are taking place in the humanities and social sciences, with databases of statistical data (Statistics Netherlands), case law and text corpora in the making. The availability of these databases and new ICT tools (both for qualitative and for quantitative methods) has made analyses possible in these fields which in the past would have been impossible or impracticable. A database containing information from the population censuses and occupational censuses will be a resource of comparable value to historians.

More specifically, the population censuses and the occupational censuses are perfectly consistent with the NIWI's stated priorities. These take close account of a) the research value and the value to the academic community of the collection to be digitised, and b) the potential for expanding the NIWI's expertise. Examples of digitisation projects (including electronic publication projects) in the field of social and economic history and with a clear research character include:

– National Accounts of the Netherlands 1800-1913 (published electronically on the Internet in co-operation with the Posthumus Institute)
– Strikes in the Netherlands 1850-1995 (in co-operation with a PhD student and the IISH)
– Historical Database of Dutch Municipalities (HDNG) *c*. 1800-1990 (in co-operation with the universities of Nijmegen and Amsterdam and Statistics Netherlands)
– Digitisation of historical climatological data from ships' logs and land-based sources (in co-operation with the KNMI, the Royal Dutch Meteorological Institute)
– Electronic publication of the 1795-1971 censuses (in image form only, except for the 1899 census).

This brief list shows that virtually all the NIWI's digitisation projects are in collaboration with or commissioned by third parties. Although NIWI policy is not aimed primarily at digitising its own collections, it does plan to make parts of the historical collections available digitally, and it will do this in direct consultation with researchers. One project is now underway in which the NIWI co-operates with researchers to digitise selected holdings from its collection. It also disseminates information through publications, courses, workshops, summer schools and congresses. At the international level too, the NIWI is a valued partner in such projects as the European Visual Archive (whose participants also include the London Metropolitan Archives and the Antwerp City Archives) and SEPIA (Safeguarding European Photographic Images for Access).

8.3 The HSN and the IISH

Long-term strategy
The HSN's strategy comprises three elements. First, it aims to promote maximum use of the existing dataset. Second, it aims to extend this dataset by initiating its own projects and by organising joint projects. Third, it aims to preserve current expertise and to expand it where possible. This expertise is focused on two areas: a) historical knowledge relating to the sources used, and b) knowledge relating to the computerisation of data. This expertise could also be described as involving the systematic and professional input, documentation, archiving and dissemination of historical microdata.

In setting up projects the HSN is guided by the expertise available in the fields in question. This expertise is accessed through the HSN steering group. A comprehensive description of the HSN's data-collection strategy can be found in HSN working paper no. 37, which sets out the organisation's longer-term perspective.

Use of data
In the coming decades, the HSN data are expected to be a unique, authoritative source for historically related projects in a wide range of disciplines. These projects can utilise the HSN data and infrastructure in three ways.

First, historical research questions in the fields of demography, sociology, epidemiology, social economics and human geography can be addressed using the available data. The data collected so far have been used in a total of seventeen NWO-funded projects (seven postdoctoral researchs and five PhD researchers; see the summary in Table 4, appendix B1). The data on the Province of Utrecht have already been analysed in a pilot project by a working group specially formed for that purpose. One aim was to illustrate the value of this material for a broader group of potential users. The working group's findings are reported in Mandemakers & Boonstra 1995 [The Life Course of the Population of Utrecht in the 19th century]. The material on Zeeland and Friesland was the subject of symposia held in March 1998 and February 2001 (see the HSN annual report, appendix F for a complete list of studies where HSN data have been used). The data from the HSN sample are likely to form the basis of many more analyses addressing these types of questions in future. In view of the size and nationwide scope of the completed sample, interregional and international comparisons will also become increasingly feasible. The international

workshop organised by the HSN in May 2001 agreed a standard for how to deal with this type of database (Amsterdam protocol, *Best practices with large databases on historical populations*), and the various papers presented also gave valuable insights into the wealth of the material.

Second, the HSN sample can form the basis for more specifically defined datasets, created by supplementing HSN data by data from other historical sources (for example church registers, tax records, spatial contexts). Coupling to other datasets is, however, subject to certain conditions. If further research is to be based on the HSN database, the researchers must adopt its data structures, and their data must generally meet the quality criteria set by the HSN for its own database. Hilda Bras's PhD study of domestic servants in Zeeland is one example of such research. Bras has reconstructed the life course of women and their sisters (if any) and investigated how their life chances depended on the social background of their parents, their migration histories and the social status of their employers. Her data include information on each subject's migration history, any marriage certificates, family situation at birth and at start of occupational career, and incomes of fathers, employers, and husbands.

Third, the HSN offers not just data, but also software for inputting and processing the data. Use of this software not only ensures standardised data input but also saves researchers vast amounts of time and money, since they no longer have to design and implement the software they need.

Policy

The HSN Foundation's policy has always aimed at creating a database that would benefit a broad range of research questions and disciplines. It has drawn on the finest expertise in the Netherlands and abroad. Some such contacts go beyond consultations and are institutionalised, with collaborating organisations represented on the HSN's board or Advisory Council. An account of all HSN activities appears in its annual report.

The following points are particularly worth highlighting:

– The creation of the HSN was the initiative of scholars from a number of disciplines, representing both the humanities and the social sciences. This interdisciplinary origin is reflected in the composition of the HSN board (see appendix C1)
– The manual archive work entailed in collecting the historical data is carried out by well-trained, qualified staff, supervised by professional historians with extensive experience of this type of database and first-rate expertise in the field of historical computing.
– The HSN has a multidisciplinary Advisory Council, made up of professors and eminent researchers from related fields (see Appendix C2).

The relationship between the IISH and the HSN Foundation

The International Institute of Social History (IISH) provides accommodation for the HSN and its activities, and also bears the associated costs. The contractual relationship between the IISH and the HSN recognises the autonomy of the HSN board in managing the data. In conducting the current projects, the HSN consults regularly with the IISH within the HSN steering group.

The IISH has guaranteed to provide 0.5 FTE for an indefinite period for co-ordination work. The availability of a half-time HSN co-ordinator, and the accommodation and other facilities provided by the IISH, will ensure that the HSN data remain accessible for national and international use in the long term too.

LITERATURE

Delger, H. & J. Kok. 1998. Bridegrooms and biases: a critical look at the study of intergenerational mobility on the basis of marriage certificates. *Historical Methods.* 31: 113-121.

Liefbroer, A.C. & P A. Dykstra. 2000. *Levenslopen in verandering. Een studie naar ontwikkelingen in de levenslopen van Nederlanders geboren tussen 1900 en 1970.* Den Haag: Sdu Uitgevers

Kelly Hall, P., B. McCaa, & G. Thorvaldsen (eds.). 2000. *Handbook of International Historical Microdata for Population Research.* Minneapolis: Minnesota Population Center.

Knotter, A. & A.C. Meijer (eds.). 1995. De gemeentelijke bevolkingsregisters, 1850-1920. pp. 77-116 in *Broncommentaren* 2. Den Haag: Instituut voor Nederlandse Geschiedenis.

Mandemakers, K. & O. Boonstra (eds.). 1995. *De levensloop van de Utrechtse bevolking in de 19e eeuw.* Assen: Van Gorcum.

Viskil, E. 1999. *Een digitale bibliotheek voor de geesteswetenschappen*. Beleidsnota gebiedsbestuur geesteswetenschappen NWO.

Van Leeuwen, M. & I. Maas. 1997. Social Mobility in a Dutch Province, Utrecht 1850-1940. *Journal of Social History.* 30: 619-644.

LIST OF APPENDICES

A1 Publications by Members of the Academic Supervisory Board
A2 Publications Executive Management.

B1 Budget Life Courses
B2 Budget Census Digitisation

C1 Members HSN Board
C2 HSN Advisory Council
C3 Privacy Statute
C4 List of Variables HSN-database

D Research Programme Life Courses in Context.

APPENDIX A RELEVANT PUBLICATIONS

A1 Members Academic Supervisory Board

Engelen, Th.L.M. & P.M.M. Klep. 2001. Wonen in Nederland op de drempel van de twintigste eeuw. rond 1900. pp. 113-130 in J.G.S.J. van Maarseveen & P.K. Doorn (red.). *Nederland een eeuw geleden geteld.* Amsterdam: Stichting beheer IISG.

Engelen, Th.L.M. 1999. The development of rgional patterns of nuptiality in 20th-century Europe. pp. 273-288 in I. Devos & L. Kennedy (eds.). *Marriage and Rural Economy. Western Europe since 1400.* Turnhout.

Engelen, Th.L.M. 1997. The fertility decline in the Dutch province of Limburg 1880-1960. On understanding historical actors in constrained environment. *The history of the Family. An International Quarterly* 2: 405-424.

Baud, M. & Th. Engelen. (red.). 1994. *Samen wonen, samen werken? Vijf essays over de geschiedenis van arbeid en gezin*. Hilversum: Verloren.

Engelen, Th.L.M. & J.H.A. Hillebrand. 1986. Fertility and nuptiality in the Netherlands 1850-1960. *Population Studies* 40: 487-503.

Treiman, D.J. & H.B.G. Ganzeboom. 2000. The Fourth Generation of Comparative Stratification Research. pp. 123-150 in: S. Quah & A.Sales (eds.). *The International Handbook of Sociology.* Sage.

Ganzeboom, H.B.G. & W.C. Ultee (red.). 1996. *De sociale segmentatie van Nederland in 2015.* Den Haag: Sdu Uitgevers. Voorstudie en achtergrond Wetenschappelijke Raad voor het Regeringsbeleid V96.

Ganzeboom, H.B.G. & R. Luijkx. 1995. Intergenerationele beroepsmobiliteit in Nederland: patronen en historische veranderingen. pp. 14-30 in J. Dronkers & W.C. Ultee (eds.). *Verschuivende ongelijkheid in Nederland. Sociale gelaagdheid en mobiliteit*. Assen: Van Gorcum.

Graaf, P. de & H.B.G. Ganzeboom. 1993. Family Background and Educational Attainment in the Netherlands of 1891-1960 Birth Cohorts. pp. 75-100 in: Shavit, Y. & H.-P. Blossfeld (eds.). *Persistent Inequality. Changing Educational Attainment in Thirteen Countries.* Boulder CO: Westview Press.

Ganzeboom, H.B.G. 1993. *Rondom Data. Over het empirische onderdeel van de empirisch-theoretische sociologie*. Oratie Universiteit Utrecht.

Heerma van Voss, L. 2000. De rode dreiging en het verzuilde antwoord. pp. 115-132 in J.H.C. Blom & J. Talsma (eds.). *De verzuiling voorbij. Godsdienst, stand en natie in de lange negentiende eeuw.* Amsterdam: Het Spinhuis.

Davies, S., C.J. Davis, D. de Vries, L. Heerma van Voss, L. Hesselink & K. Weinhauer (eds.). 2000. *Dock Workers. International Explorations in Comparative Labour History, 1790-1970.* Aldershot: Ashgate. 2 vols.

Heerma van Voss, L. & F. Vermeulen. 2000. Back to the country. Living Strategies of Twente Textile Workers ca. 1920. *Working paper HSN.*

Heerma van Voss, L. 1995. The Netherlands. pp. 39-70 in S. Berger & D. Broughton (eds.). *The force of labour. The Western European labour movement and the working class in the twentieth century*. Oxford/Washington D.C.: Berg.

Heerma van Voss, L. 1994. *De doodsklok van den goeden ouden tijd. De achturendag in de jaren twintig*, Amsterdam: Stichting beheer IISG.

Knippenberg, H. & H. van der Wusten. 2001. De zuilen, hun lokale manifestaties en hun restanten in vergelijkend perspectief. in C. van Eijl, L. Heerma van Voss & P. de Rooy (red.). *Het sociale Nederland, contouren van de twintigste eeuw.* Amsterdam: Het Spinhuis/IISG.

Knippenberg, H. 2001.Polarisatie en versnippering: kerk en godsdienst rond 1900. pp. 131-157 in J.G.S.J. van Maarseveen & P.K. Doorn (red.). *Nederland een eeuw geleden geteld. Een terugblik op de samenleving rond 1900*. Amsterdam/Voorburg: IISG/NIWI/CBS.

Knippenberg, H. 1999. National integration and growing ethnicity among Dutch Roman Catholics and Protestants in the nineteenth and early twentieth century. The impact of education. *The Netherlands' Journal of Social Sciences.* 35: 37-52.

Knippenberg, H. 1992. *De religieuze kaart van Nederland. Omvang en geografische spreiding van de*

godsdienstige gezindten vanaf de Reformatie tot heden. Assen/Maastricht: Van Gorcum.
Knippenberg, H. & B.C. de Pater. 1988. *De eenwording van Nederland. Schaalvergroting en integratie sinds 1800.* Nijmegen: SUN (2e herziene druk 1990, 3e druk 1992, 4e druk 1997, 5e druk 2000).

Kok, J. 2000. Transities en trajecten. De levensloopbenadering in de sociale geschiedenis. *ijdschrift voor Sociale Geschiedenis* 26: 309-329.
Kok. J., 1999. Migratie als gezinsstrategie in midden-Nederland, 1850-1940. pp. 89-156 in J. Kok, A. Knotter, R. Paping. & E. Vanhaute (red.). *Levensloop en levenslot. Arbeidsstrategieën van gezinnen in de negentiende en twintigste eeuw.*Groningen/Wageningen.
Kok, J. & H. Delger. 1998. Success or selection? The effect of migration on occupational mobility in a Dutch province, 1840-1950. *Histoire & Mesure* 13: 289-322.
Kok, J. 1997. Youth labor migration and its family setting, The Netherlands 1850-1940. *History of the Family. An International Quarterly.* 2: 507-526.
Kok, J, 1994. Revealing family strategies using life course analysis. *Economic and social history in the Netherlands* 6: 109-127.

Leeuwen, M. van & I. Maas. *fortcoming* 2002. Partner choice and homogamy in the nineteenth century: Was there a sexual revolution in Europe?' *Journal of Social History*.
Leeuwen, M. van 2000. *Logic of Charity. A Simple Model Applied to Amsterdam 1800-1850.* London: Macmillan.
Leeuwen, M.H.D. van 2000. *De rijke Republiek. Gilden, assuradeurs en armenzorg 1500-1800; en De eenheidsstaat: onderlinges, armenzorg en commerciële verzekeraars 1800-1890.* Delen 1 en 2 van J. van Gerwen, J. & M.H.D. van Leeuwen (red.). *Zoeken naar Zekerheid. Risico's, preventie, verzekeringen en andere zekerheidsregelingen in Nederland 1500-2000.* Amsterdam: NEHA.
Leeuwen, M.van & J.A. Faber. 1991. Les assistés catholiques d'Amsterdam, 1750-1850. Un example d'application particulière de la methode de reconstitution des familles. *Annales de Démographie Historique*: 239-253..
Leeuwen, M. van & I. Maas. 1990. De openheid van de Amsterdamse elite tijdens de industrialisatie: loglineare analyse van carrièremobiliteit, intergenerationele mobiliteit en sociaal netwerk. *Mens en Maatschappij* 65: 5-26.

Lucassen, J. & L. Lucassen (eds.). 1997. *Migration, Migration History, History: Old Paradigms and New Perspectives* Bern, etc.: Peter Lang [paperback 1999].
Lucassen, J. & R. Penninx. 1997. *Newcomers. Immigrants and their Descendants in the Netherlands 1550-1995*. Amsterdam: Het Spinhuis.
Lucassen, J. 1996. The North Sea: a crossroad for migrants? pp. 168-184 in: J. Roding & L. Heerma van Voss (eds.). *The North Sea and Culture (1550-1800).* Hilversum: Verloren.
't Hart, M., J. Lucassen & H. Schmal (red.). 1996. *Nieuwe Nederlanders. Vestiging van migranten door de eeuwen heen*. Amsterdam: IISG/SISWO.
Lucassen, J. 1995. Emigration to the Dutch Colonies and the USA. pp. 21-27 in R. Cohen, (ed.). *The Cambridge Survey of World Migration.* Cambridge [Section 2 European Colonization and Settlement]: 21-27.

Maarseveen, J.G.S.J. van & P.K. Doorn (red.). 2001. *Nederland een eeuw geleden geteld: terugblik op de samenleving rond 1900.* Amsterdam: Stichting beheer IISG.
Doorn, P.K., J.K. Jonker, J.G.S.J.van Maarseveen & T. Vreugdenhil. 1999. *Publicaties Volkstellingen 1795-1971.* CD-Rom Amsterdam: Stichting beheer IISG.
Maarseveen, J.G.S.J. van & M.B.G. Gircour (eds.). 1999. *A Century of Statistics. Counting, accounting and recounting in the Netherlands*. Amsterdam: Stichting beheer IISG. [Vertaling van *Een eeuw statistieken*. Hierin ook een tweetal bijdragen van J.G.S.J. van Maarseveen, waarvan een over de geschiedenis van de volkstellingen].
Maarseveen, J.G.S.J. van, M.B.G.Gircour & R.Schreijnders. (eds.). 1999. *A Century rounded up. Reflections on the history of the Central Bureau of Statistics in the Netherlands*. Amsterdam: Stichting beheer IISG. [Vertaling van *Een eeuw welgeteld*. Hierin ook een drietal bijdragen van J.G.S.J.van Maarseveen waaronder over ´statistical data collection´).
Maarseveen, J.G.S.J. van. 1990-1993. *Briefwisseling van Nicolaas Gerard Pierson 1839-1909.* Amsterdam. 4

vol. (bronnepublicatie).

Maas, Ineke & R. A. Settersten, Jr. 1999. Military service during wartime: its effects on men's occupational trajectories and later economic well-being. *European Sociological Review.* 15: 81-100.

Leeuwen, Marco H.D. & I. Maas. 1997. Social Mobility in a Dutch Province, Utrecht 1850- 1940. *Journal of Social History*. 30: 619-644.

Aschaffenburg, K. & I. Maas. 1997. Cultural and educational careers: The dynamics of social reproduction. *American Sociological Review.* 62: 573-587.

Leeuwen, M.H.D. van & I. Maas. 1996. Long-term social mobility: research agenda and a case study (Berlin, 1825-1957). *Continuity and Change*. 11: 399-433.

Leeuwen, M. H.D. van & I. Maas. 1995. Groeiende openheid van de Nederlandse samenleving: een nieuw fenomeen of een lange trend? Intergenerationele, huwelijks- en carrièremobiliteit in de provincie Utrecht, 1850-1940. *Mens en Maatschappij.* 70: 321-333

Poppel, F. van & I. Joung. 2001. Long-term trends in marital status mortality differences in the Netherlands 1850-1970. *Journal of Biosocial Science* 33: 279-303.

Poppel, F. van, A. Liefbroer, J. Vermunt & W. Smeenk. 2001. Love, necessity and opportunity: Changing patterns of marital age homogamy in the Netherlands, 1850-1993. *Population Studies.* 55: 1-13.

Poppel, F. van. 2000. Children in one-parent families: Survival as an indicator of the role of the parents. *Journal of Family History* 25: 269-290.

Poppel, F. van. 1998. Nineteenth-century remarriage patterns in the Netherlands *Journal of Interdisciplinary History*. 28: 343-383.

Joung, I.M.A., H.D. van de Mheen, K. Stronks, F. W. A. van Poppel & J.P. Mackenbach. 1998. A longitudinal study of health selection in marital transitions. *Social science and medicine* 46: 425-435.

Post, W., F. van Poppel, E. van Imhoff & E. Kruse. 1997. Reconstructing the extended kin-network in the Netherlands with genealogical data: problems, methods and results. *Population Studies* 51: 263-278.

Wals, H., forthcoming. The Trade Union as Survival Strategy. The Case of Amsterdam Construction Workers in the First Quarter of the Twentieth Century. in J. Kok (ed.). *Rebellious families. Household strategies and collective action in the nineteenth and twentieth centuries*. Oxford.

Wals, H. 2001. *Makers en Stakers. Amsterdamse bouwvakarbeiders en hun bestaansstrategieën in het eerste kwart van de twintigste eeuw.* Amsterdam: Stichting beheer IISG.

A2 Executive Management.

Maarseveen, J. van & P. Doorn (red.). 2001. *Nederland een eeuw geleden geteld: terugblik op de samenleving rond 1900.* Amsterdam: Stichting beheer IISG.

Doorn, P. 1998. The Old and the Beautiful: a soap opera about misunderstandings between historians and models. *VGI Cahier* 10: 9-30.

Boonstra, O.W.A., P K. Doorn & F.M.M. Hendrickx. 1990. *Voortgezette statistiek voor historici.* Muiderberg: Coutinho

Breure, L., O.W.A. Boonstra & P.K. Doorn (red.). 1990. *Historische informatiekunde* Hilversum: Verloren.

Doorn, P.K. 1989. *Social structure and spatial mobility: composition and dynamics of the Dutch labour force* . Dissertatie Universiteit Utrecht.

Mandemakers, C.A. 2001. De sociale structuur in Nederland rond 1900. De samenleving in het perspectief van de modernisering 1850-1990. pp. 185-208 in J.G.S.J. van Maarseveen & P.K. Doorn (red.). *Nederland een eeuw geleden geteld.* Amsterdam: Stichting beheer IISG

Mandemakers, K. 2000. The Netherlands. Historical Sample of the Netherlands. p. 149-177 in P. Kelly Hall, R. McCaa & G. Thorvaldsen (eds.). *Handbook of International Historical Microdata for Population Research.* Minneapolis: Minnesota Population Center. .

Van Poppel, F. & K. Mandemakers. 1997. Differential infant and child mortality in the Netherlands 1812-1912. p. 276-300 in: A. Bideau, B. Desjardins & H. Pérez-Brignoli (eds.). *Infant and child mortality in the past.*

Oxford: Clarendon Press.
Mandemakers, C.A. 1996. *Gymnasiaal en middelbaar onderwijs: ontwikkeling, structuur, sociale achtergrond en schoolprestaties, Nederland, ca. 1800 - 1968.* Proefschrift Erasmus Universiteit Rotterdam.
Mandemakers, K. & O. Boonstra (red.). 1995. *De levensloop van de Utrechtse bevolking in de 19e eeuw.* Assen: Van Gorcum.

APPENDIX B1 BUDGET FOR LIFE COURSES

Projected scale
The budget includes a) the costs of entering population register data for subjects born between 1863 and 1922 and b) the costs of entering the data from marriage certificates and death certificates of these subjects as far as these are not already included in the HSN database.

Table 1 Workload: inputting population register data

Birth cohorts	Subjects	Already input at start of project	Not yet input	Workload per day	No. of working days
Initial registrations in population registers					
1863-1874	10500	4500	6000	12	500
1875-1899	18500	7000	11500	15	767
1900-1922	12000	4000	8000	15	533
Total	41000	15500	25500		1800
Subsequent registrations in population registers (up to 1940)					
1863-1874 [a]	8000	400	7600	3	2533
1875-1899 [b]	15000	400	14600	45	3244
1900-1922 [c]	11000	200	10800	6	1800
Total	34000	1000	33000		7577
	Days	Years [d]			
Total workload	9377	49			

a 25% fewer subsequent registrations because of infant and childhood mortality
b 20% fewer subsequent registrations because of infant and childhood mortality
c 10% fewer subsequent registrations because of infant and childhood mortality
d Number of working days a year is 190 (1330 workable hours a year)

Table 2 Workload: inputting certificates of birth, marriage and death, and personal record cards

	Subjects	Already input at start of project	Not yet input	Workload per day	No. of working days
Marriage certificates	32000	14000	18000	15	1200
Death certificates	18000	8000	10000	20	500
Personal record cards of spouses [a]	11000	0	11000	35	314
Personal record lists of spouses [a,b]	3000	0	3000	0 [b]	0
					2014
	Days	Years [c]			
Total workload	2014	11			

a Of female subjects
b Personal record lists will be processed fully automatically
c Number of working days a year is 190 (1330 workable hours a year)

Costs
The total workload of Table 1 and 2 expressed in years is 49+11=60. All the work can be completed in five years using a team of 12 data-entry workers. The project will be carried out under the auspices of the International Institute of Social History (IISH) in Amsterdam.

The following budget (Table 3) covers the material costs of obtaining, inputting and managing the data to be collated.

Table 3 Budget: inputting life-course data, at May 2001 prices (EUR '000)

	Fte	*Yearly cost*	*Total Per Year*	*Entire Period*
PERSONNEL COSTS				
1. Programme management (scale 12)	11	70	77	385
2. Database administrator (scale 9)	1	44	44	220
3. Programmer (scale 9)	1	44	44	220
4. Supervisory historians (scale 10)	2	50	100	500
5. Archive work by historians (scale 8)	10	39	390	1950
Total personnel costs			655	3275
MATERIAL COSTS				
1. Purchase and maintenance of office hardware (workstation,PCs, printers)			7	35
2. Purchase and maintenance of electronic notebooks	14	4	6	30
3. Purchase and updating of software			5	25
4. Costs of postage, statutory fees			23	115
5. Costs of personal record cards, personal record lists			11	55
6. Travel costs (domestic)	13	27	35	175
7. Travel and accommodation costs (international)			6	30
Total material costs			93	465
Total personnel + material costs			748	3740
TOTAL COSTS			748	3740
REVENUES			748	3740
Programme for NWO Large Investments			638	3190
Matching funds from joint projects			70	50
Matching funds (IISH 0.6 FTE scale 12)			40	200

Notes to the personnel costs
General co-ordination: On the basis of experience gained over the past four years, we estimate the time required for overall co-ordination, liaison with the HSN board and the IISH management, information analysis, software development, supervision of data-entry work (including supervising the database administrator, the programmer and the historians carrying out the work), plus the time required for promoting the HSN project in the Netherlands and abroad, etc., at 1.1 FTE in scale 12.
Database administrator: The principal task of the database administrator is to process and monitor the data being input. He or she is also responsible for the preparatory work at the various archives. The database administrator maintains the software necessary to carry out these tasks.

Programmer: To make the data accessible to researchers through data releases and direct Internet access, we will need to develop and maintain software. In addition, database-management and data-entry programs must be developed and maintained. (Programs for inputting marriage certificates, personal record cards and population registers are already available.)
Archive work by historians: In the intended research design, the data-entry work will be carried out by qualified historians. This is unlike the original HSN design, where data-entry workers were recruited largely through the Jobseekers Employment (WIW) Scheme. The current state of the labour market makes it increasingly difficult to find suitable scheme participants, while the current proposal also necessitates rigid production planning and guaranteed high labour productivity. No compulsory redundancies will occur, however, and as far as possible use will be made of the capacity still available. In terms of costs, three scheme participants (incl. supervision) are roughly equivalent to one historian (the productivity of one historian will not be much lower than that of three participants though). The number of staff required is twelve, for a period of five years (see Tables 1 and 2); two of these will have a more supervisory role.

Notes to the material costs
Office hardware: Total annual hardware costs are _ 7000. This comprises annual depreciation of the server (_ 2000) and four high-performance office PCs (per computer _ 750 a year). The printing facilities will require annual depreciation costs of _ 500, plus paper and toner at _ 1500.
Electronic notebooks: Project staff will be equipped with electronic notebooks. The cost of depreciation is currently _ 400 per notebook per year. The price of notebooks is not expected to decrease any further. An additional notebook, for ambulant use (as well as data entry), is included in the budget.
Software: This relates principally to the costs of maintaining and developing the current database software. In addition, many basically one-off purchases of auxiliary software and updates will be necessary. These costs are estimated at _ 5000 a year on average.
Costs of postage, statutory fees: In addition to the personal record cards, copies of *c*. 5000 marriage certificates will have to be ordered from the municipalities (applying to marriages after 1940). A fee of _ 8 per certificate is charged, giving a total of _ 40,000. It will not always be possible for us to work at the population registers ourselves. Particularly in the case of small or remote archives, it will be more efficient to commission others to do the work (around 20% of cases, i.e. 6500 applications). These costs are estimated at _ 75,000 (average of _ 11.50 per application, incl. postage). Total costs for this item are thus _ 115,000 (annual average of _ 23,000).
Costs of personal record cards: The average cost per personal record card is _ 4; the average cost per personal record list is slightly less (depending somewhat on how often a request is made). The number of cards required is 11,000, and the number of lists 3000. Total cost _ 55,000.
Travel costs (domestic): The work will be carried out in archives across the country. Project staff will need a public transport season ticket, costing _ 2700 a year. An additional _ 2700 is included for miscellaneous travel costs.
Travel and accommodation costs (international): To attract foreign matching projects and promote the international use of the dataset, we will need to attend conferences and make working visits to academics interested in collaborating with the HSN. To develop the HSN as an organisation, we will need to build and maintain contacts with foreign sister organisations. These costs are estimated at _ 6000 a year.
Office accommodation and organisational support: The IISH has passed on neither the costs of office accommodation nor the costs of the support it gives to the HSN. In this way it subsidises the current projects. The present application therefore makes no provision for these costs.

Financial contributions from other sources (matching funds)
Matching funds will be received from the IISH and from future joint projects with other organisations. The IISH will contribute 0.6 FTE for co-ordination and management. As the overview below indicates, the HSN is capable of securing a regular flow of project funds; these have averaged _ 40,000 a year over the past nine years. This figure can probably be increased to _ 70,000 a year.

The HSN is an attractive partner in projects requiring social-historical and historical-demographic data at the individual level. The HSN has participated, or is currently participating, in more than ten such projects (see Table 4). These include joint studies with epidemiologists from the University of Nijmegen (research on the effects of the timing of ovum fertilisation on the later fecundity of a female born from that egg cell); with historical demographers from the University of Nijmegen (research on regional differences in demographic behaviour); with the Posthumus Institute (research on occupational mobility, migration and the labour strategies of families); with the Interuniversity Center for Social Science Theory and Methodology (ICS) (research on geographical and social mobility of female domestic servants), with the Netherlands Interdisciplinary Demographic Institute (NIDI) (research on childhood mortality), and with migration historians (University of Amsterdam and the Centre for the History of Migrants (CGM)).

On the basis of past experience, it is highly likely that the HSN will continue to attract new project funds. There are several further ongoing projects in which the HSN collaborates with other researchers by providing software and software support. In exchange, those projects enter their data according to HSN guidelines, which means the HSN database acquires new data easily and in a compatible format. For an overview, see table 4. This overview not only includes funds *going directly to the HSN (about 3,5 million guilders till now)*, but the various participants have also invested *around 3,5 million additional guilders* in research associated with the HSN.

Table 4 List of subsidised HSN-projects.

Project	Period	Funding HSN			Total costs project (including research)
		Direct	Indirect (time)	Total	
HSN Pilot project , Province of Utrecht (Ministry of Education and Science)	1991-1993	220		220	220
On-off continuation grant (Ministry of Education and Science)	1993-1995	200		200	200
“Occupational Mobility and migration” in collaboration with the N.W. Posthumus Institute	1993-1996	50		50	103
“Reduced fecundity:, in collaboration with the University of Nijmegen (NWO: 900-561-064)	1993-1996	214	30	244	500
“Regional differences in demographic behaviour, the Netherlands”, in collaboration with the University of Nijmegen (NWO 240-129-046, foundation HSN)	1994-1997	76	50	126	400
“HSN-Zeeland research” in collaboration with the Stichting Regionale Geschiedbeoefening Zeeland (SRGZ)	1994-2001	25		25	25

Project	Period	Funding HSN			Total costs project (including research)
		Direct	Indirect (time)	Total	
Expansion of the HSN base sample to the whole of the Netherlands (programme for NWO Medium-Sized Investments)	1996-1999	1200		1200	1200
"Labour strategies of families", in collaboration with the N.W. Posthumus Institute (NWO N14-96)	1996-1998	30		30	84
"Geographical and social mobility of female domestic servants in Zeeland 1860-1920", in collaboration with the University of Utrecht (ICS)	1997-2001		70	70	300
"Textile Industry Workers in Twente", pilot research within the research programme on living strategies (IISH)	1997-2001		10	10	10
"Germans in Utrecht: a temporary minority in the 19th century" (NWO 250 290 53)	1998-2001	3		3	300
"Religion and childhood mortality in the Netherlands, 1860-1920", in collaboration with the NIDI and the Hebrew University, Jerusalem (Wellcome Trust)	1999-2001		100	100	220
Pioneer project "Settlement determinants for immigrants and their descendants in the Netherlands, 1853-1960" (NWO, University of Amsterdam and CGM)	1998-2002	400	30	430	1000
Expansion of the HSN base sample to include marriage certificates and initial registrations in the population register (Programme for NWO Medium-Sized Investments)	2000-2001	750		750	750
"Family establishment and living strategies in the West of the Netherlands (1830-1940)" (Population and Society in Taiwan and the Netherlands) (NWO 245-53-001, University of Nijmegen, Academia Sinica (Taiwan), Stanford University)	2000-2002	100	40	140	300
'Early Childhood, Social Mobility and Longevity (National Institute of Health, Washington, NIDI, Indiana University Bloomington)	2001-2004	400	25	425	1000
PhD research, "Three generations of Rotterdam Labourers" (Reaal Insurances)	2001-2005		50	50	350
Post-doc research 'Europese migratie Nederlands-Indië' (NWO-	2001-2005		75	75	322

Project	Period	Funding HSN			Total costs project (including research)
		Direct	Indirect (time)	Total	
355-53-004)					
Total		3668	480	4148	7284

APPENDIX B2 BUDGET FOR DIGITISATION OF CENSUSES

Nr	Activity	Implementation	Working months	Costs/ month	Total Costs (Euro)	Total Costs (Dfl.)
1	PREPARATORY WORK					
1.1	Selecting material	R	Finished			
1.2	Preparing material	L	Finished			
1.3	Microfilming	O	Finished			
2	TECHNICAL RESEARCH					
2.1	Researching automatic character/structure recognition	R	6	4,800	28,800	63,467
2.2	Researching media conversion	R/L	Finished			
3	RESEARCHING/PREPARING DATA STORAGE					
3.1	Document analysis/information analysis	R	3	4,800	14,400	31,733
3.2	Data modelling	R/A	Finished			
3.3	Preparing/analysing metadata	R	2	4,400	8,800	19,393
3.4	Preparing conversion of OCR output to database	R	1	4,800	4,800	10,578
3.5	Preparing data entry	R/L	1	4,800	4,800	10,578
4	USERS/MARKET RESEARCH					
4.1	Research among users	R	Finished			
4.2	Market research	R	Finished			
5	MEDIA CONVERSION (IMAGING)					
5.1	Scanning microfilms	L	Finished			
5.2	Creating metadata scans	L	Finished			
6	CONTENT CONVERSION					
6.1	Optical character recognition	L	1	3,300	3,300	7,272
6.2	Content conversion by data entry	L/O	100	3,000	300,000	661,113
7	DIGITAL CONVERSION					
7.1	Analysis of current systems	R	PM			
7.2	Automated conversion and structuring routines	R/A	PM			

Nr	Activity	Implementation	Working months	Costs/ month	Total Costs (Euro)	Total Costs (Dfl.)
7.3	Manual structuring	L	PM			
8	POST PROCESSING					
8.1	Checking/correcting media conversion	L/R	Finished			
8.2	Checking/correcting metadata	L/R	Finished			
8.3	Checking/correcting OCR/data entry	L/R	6	4,400	26,400	58,178
9	IMPLEMENTING RETRIEVAL/ACCESS					
9.1	Functional analysis	R/A	3	4,800	14,400	31,733
9.2	Technical design	R/A	2	4,800	9,600	21,156
9.3	Graphic design	R/A/O	Finished			
9.4	Programming	A/R/O	4	4,800	19,200	42,311
10	ANALYSIS					
10.1	Source criticism, classifications, categories	R	8	4,800	38,400	84,622
10.2	Temporal links/comparisons	R	6	4,800	28,800	63,467
10.3	Contacts with foreign projects	R	PM			
10.4	Substantive/methodological research	R	12	5,500	66,000	145,445
10.5	Publications and presentations	R	6	6,400	38,400	84,622
11	TESTING AND INSTALLATION, PRODUCTION					
11.1	Designing test procedures	R/A	1	4,800	4,800	10,578
11.2	Implementing user tests	R	1	4,800	4,800	10,578
11.3	Debugging	A/R/O	2	4,800	9,600	21,156
11.4	Production of digital storage medium (CD-ROM)	O	1		10,000	22,037
11.5	User instructions		PM			
11.6	Installation onto production system	S	1	4,800	4,800	10,578
12	PROJECT MANAGEMENT AND ORGANISATION					
12.1	Co-ordination and consultation	R	3	4,800	14,400	31,733
122	Quality management	R/L	3	4,800	14,400	31,733
123	Project management	R	3	6,400	19,200	42,311
13	DIGITAL ARCHIVING					
13.1	Conversion to software-independent storage format	R/A	1	4,400	4,400	9,696
13.2	Documentation	R	3	4,400	13,200	29,089
13.3	Storage	R/S	1	4,400	4,400	9,696
14	OPERATION, MAINTENANCE AND SUPPORT					
14.1	Helpdesk	R/S	PM			
14.2	System maintenance	R	PM			
14.3	Web hosting	R/S	PM			
	SUBTOTAL PERSONNEL COSTS (incl. work outsourced)		**181**		**710,100**	**1,564,854**

R = Academic staff

L = Digitization & Library services

Nr	Activity	Implementation	Working months	Costs/ month	Total Costs (Euro)	Total Costs (Dfl.)
	A = Application design					
	S = IT Support					
	O = Outsourcing					
	Material costs					
	Server costs, storage, etc.				20,000	44,074
	PCs, electronic notebooks, printers, etc.				25,000	55,093
	Software + digitization map boundaries				55,000	121,204
	Travel and accommodation expenses				15,000	33,056
	Publishing costs				10,000	22,037
	SUBTOTAL MATERIAL COSTS				**125,000**	**275,464**
	TOTAL				**835,100**	**1,840,318**
	REVENUES				835,100	1,840,318
	Programma for NWO large investments				495,000	1,090,836
	Matching funds from joint projects				180,000	396.668
	Matching funds NIWI				160,100	352.814

Notes to activity groups

General
The budget is based on the standard activity scheme as used in digitization projects at NIWI-KNAW. For the general overview, activities that already have been finished in earlier projects are included in the scheme. Some activities that do not apply to this particular project are marked as PM.

1. Preparatory work
The selection of the materials to be digitized and its pre-processing, including microfilming, have already been carried out in an earlier project and in the preparation of this proposal.

2. Technical research
Optical character recognition of tables is not a straightforward, mainly because of the structure of the tables in (hierarchical) rows and columns. Additional problems are related to the lines in the tables, confusion of some alphabetic characters with numbers (I and 1, O and 0), separation of row and column labels from table content, etc. Earlier research has pointed out that many of the problems can be solved, although the OCR software presently available (adapted and extended at NIWI-KNAW for census tables) can not be used economically for large tables. Within the present project, it is worthwile to test the latest available software and to investigate its applicability for a variety of census tables. Some programming will be required for adapting commercially available OCR software to the specific needs of the project.

3. Researching/preparing data storage
Although the software to be used for data publishing is readily available (CBS-StatLine), the census tables need to be analysed carefully in order to be able to convert the data to the software format in such a way, that the original source structure can be reconstructed by the end user. The documentation (meta data) needed to create StatLine tables also has to be studied and prepared. A detailed data entry instruction must be made.

4. Users/market research
In earlier projects a user research has been carried out to establish the demands of the users and the nature of the 'market' for digital census information. This information is available and it is not useful to repeat such a research.

5. Media conversion (imaging)
All published census books have already been digitized.

6. Content conversion
The content conversion, in which the images are converted to data that can be analysed, is the most time consuming activity of the project. The budget includes the costs for the full digitisation of the published population, labour and housing censuses between 1859 and 1947, building upon work carried out in earlier projects.

Census year	Pages	Volumes	Labour months	Conversion ready?
1859	1186	3	7,5	Yes
1869	888	3	6,5	No
1879	2268	12	11,2	No
1889	10097	25	50,0	No
1899	10319	27	51,1	Yes
1909	3572	7	21,4	No
1919	175	1	0,7	No
1920	1903	10	16,0	No
1930	2392	15	25,8	Yes
1947	1430	15	72	No
1859-1947	**35652**	**133**	**2092**	
Conversion ready	15319	60	96,1	
To be Digitized	**20333**	**73**	**1131**	

Of each census year, estimates have been made of the average information content per page of the published volumes. The average number of characters per page varies between 1400 (in 1919) and 3600 (in 1930); over the whole period there are about 2100 characters per page. The information content can be subdivided into text (introductory volumes, methodological explications, etc.), images (maps, graphs, etc.) and tables. The tables make up the majority of the information content and can be subdivided into:

- row labels
- column labels
- cell contents (numbers)
- empty cells
- footnotes

The texts and smaller tables will be digitized by optical character recognition (OCR); all calculations are based on manual data entry, although wherever effective OCR will be used. The calculations in the table are based on manual data entry. It is cautiously estimated that OCR will be applied to 5-10% of the total number of pages, by which about half a year of manual data entry time can be saved.

Over the whole, about one third of the characters on the pages with tables consists of labels; the proportion of empty cells varies considerably. For labels that are repeated, intelligent, time-saving data-entry routines can be developed. Normally, an empty cell will also require a keyboard strike, but where they form empty rows or columns time can be saved. In order to reduce the number of data entry errors to an acceptable minimum, data will be entered twice. The number of characters to be input per hour is *c.* 7000.

For census years in which the data is already electronically available the published census volumes will not be keyed in. In total, about 20,000 pages of published censuses (39 million characters) will be converted, for which *c.* 100 months of data entry time are needed.

7. Digital conversion
The digital files for which the data entry has already been finished are available in a format that can be converted to StatLine directly.

8. Post processing
The post processing consists of the checking and correction of the digital files, both from the data entry and from the OCR-process. The checking will be supported by automated routines, based on comparing given totals in the source with totals that are calculated from the cell values. Errors found need to be inspected and the correction needs to be done manually. Two main types of errors must be distinguished: errors already present in the source and data-entry errors. Data-entry errors will be corrected. General rules to decide how source errors can be corrected will be made. Source errors that cannot be solved will be noted separately.

9. Implementing retrieval/access
The document analysis in activity group 3 will be supplemented with detailed functional requirements for retrieval and access to the tables in StatLine. A technical design of how the tables will be represented in StatLine will be made. After this, routines must be programmed in order to import the data tables into StatLine.

10. Analysis
The documentation meta-data that are first studied in activity group 3 need further analysis and processing in order to give the end user the maximum benefit of the data. The classifications of the 19th and early 20th century are often lacking or imperfect, and there are certainly problems to compare classifications over time. It will be investigated in how far in in which ways these problems can be solved, although the project will not enforce standard classification schemes on the data. Which comparisons to make and how to do it is the responsibility of the researcher. Approaches used in other (foreign) project will be taken into consideration.
The technical research will concentrate on methods to extract (synthetic) individual records from aggregate tables, to investigate which possibilities exist to link the census tables to the records in the population registers, and on methods of multi-level analysis (see also research programme).

11. Testing and installation, production
As soon as the StatLine implementation has been made, extensive user tests will be designed to check whether the application meets user requirements and is error-free. Time for corrections is reserved, and the final production phase (installation on servers, CD production). Production on CD or DVD is desirable because of the size of the data sets to be manipulated. End users will be able to create tables of a considerable size, that will not be transferred easily over networks with a low band width.

12. Project management and organisation
Because various institutes and a considerable number of people are involved in the project, co-ordination, contacts, consultations and project management will require a substantial amount of time.

13. Digital archiving
The digital censuses will not only be made accessible in StatLine, they also need to stay available for future use and other software packages. Therefore, the data will be documented and archived in an application independent way at the Netherlands Historical Data Archive at NIWI-KNAW. Standard data-archiving procedures of the NHDA will be applied.

14. Operation, maintenance and support
These activities will start at the very end of the project period and are therefore not included in the project. NIWI-KNAW and CBS will ensure the availability of the data in respectively the NHDA and in StatLine.

Appendix C1 Members HSN Board

Prof. dr. H.B.G. Ganzeboom, Universiteit Utrecht, chairman
Dr. H. Knippenberg, Universiteit van Amsterdam, vice-chairman / treasurer
Prof. dr. F.W.A. van Poppel, NIDI, secretary
Dr. J. Kok, IISG, member
Dr. M.H.D. van Leeuwen, NEHA, member
Prof. dr. J.M.W.G. Lucassen, IISG, VU Amsterdam, member
Dr I. Maas, Universiteit Utrecht, member

Appendix C2 HSN Advisory Council

Chairman
Prof. dr. J. Dronkers, European University, Florence.

Members
Prof. dr. M. du Bois-Reymond, RU Leiden
Drs. R.J.P. Dekker, hoofd Wetenschappelijk Statistisch Agentschap, hoofd Data Archieven NIWI
Dr. P.K. Doorn, hoofd afdeling informatietechnologie NIWI
Prof. dr. H. van Dijk, EU Rotterdam
Prof. dr. H.F.J.M. van den Eerenbeemt, em. hoogleraar KU Brabant
Prof. dr. W.Th.M. Frijhoff, VU Amsterdam
Prof. dr. J. Goudsblom, em. hoogleraar Universiteit van Amsterdam
Prof. dr. J. Gierveld, directeur NIDI, VU Amsterdam
Prof. dr. F.C.J. Ketelaar, Universiteit van Amsterdam
Dr. A.J. Lever, directeur Centraal Bureau voor Genealogie (CBG)
Prof. dr. F. N. Stokman, RU Groningen
Prof. dr. W.C. Ultee, KU Nijmegen
Prof. dr. A. van der Woude, em. hoogleraar LU Wageningen
Prof. dr. H.H. van der Wusten, Universiteit van Amsterdam
Prof. dr. J.L. van Zanden, Universiteit Utrecht

APPENDIX C3 PRIVACY REGULATIONS HSN

Because of the new law on privacy (Wet Bescherming Persoonsgegevens) the HSN privacy statute will be renewed within short time.

Regulations governing access to data owned by the Historical Sample of the Population of the Netherlands (Stichting Historische Steekproef Nederlandse Bevolking), as adopted at its Board meeting of January 21, 1993, and as deposited with the Personal Data Protection Board (Registratiekamer persoonsregistraties), and registered there with no. O-0030426.

1. The Historical Sample of the Population of the Netherlands (Stichting Historische Steekproef Nederlandse Bevolking, hereinafter called HSN) has its registered office in Utrecht (Register of Foundations Chamber of Commerce [Stichtingenregister Kamer van Koophandel], file no. S 183871).
2. These regulations are obligatory under article 5 of HSN's Articles and the Personal Data Protection Act (Wet persoonsregistraties); (Act of December 28, 1988, Book of Statutes [Staatsblad] 665, 1988).
3. The term 'data of the HSN foundation' refers to all data collected within the bounds of the foundation's objectives. They are 'data on a sample of the Dutch population, as they may be found in certificates of births, deaths and marriages, records of the population register and other historical source material,' HSN's Articles, article 2, paragraph a.
4. A distinction is made between data available to the public without restriction (hereinafter called 'free data') and data which may be made public subject to restrictions (hereinafter called 'non-free data'). In a register, a record will be kept of a) which data are contained in the archive and b) which status these data enjoy (free or non-free).
5. Free data are all data which may be consulted in public archives (as referred to in the Archiefwet [Archives Act] 1962, Book of Statutes 313, article 1, paragraph c) without restrictions of any kind.
6. Non-free data are all data not covered by article 5 of these regulations, and collected within the bounds of the foundation's objectives (see article 3). These data are managed in accordance with article 17 of the Decree of January 2, 1990, designating those registers of persons to which the articles 19, 24 and 25 of the Personal Data Protection Act do not apply.
7. With the passage of time, the data covered by article 6 may be released in strict accordance with the rules of the archives from which these data have been drawn (see article 5). A record of this will be kept in the register mentioned in article 4.

8 The data referred to in article 6 are accessible only for the purpose of scientific research, as envisaged in article 2, paragraph b of HSN's Articles. Researchers using these data will have to sign a mandatory statement to this effect. Only after this statement is signed will the Board be authorized to release the data concerned.

9. The provisions of article 8 imply, among other things, that use of 'non-free' data is allowed exclusively on condition that the anonymity of data is guaranteed, this within the meaning of article 33 of the Personal Data Protection Act.
10. The provision of article 9 grants an exemption from meeting the requirements of thePersonal Data Protection Act as regards the provisions contained in articles 28 up to and including 32, which concern the rights of interested parties to inspect and amend personal data.
11. The Board may authorize in writing another legal entity to manage access to the data, in accordance with the above articles.

Appendix C4

OVERVIEW OF HSN DATABASE VARIABLES

This appendix lists the most important variables used in reconstructing life courses, by the source in which they are found. These variables do not pertain solely to the Research Person (RP), but also to persons associated to the RP. *In this survey, the variables are not repeated which have already been found in preceding sources and which do not change (such as the name) or which can be predicted (such as age).*

Birth certificate

Father RP	Name, age, occupation, address, signature
Mother RP	Name, age, occupation, address
Witnesses (two)	Name, age, occupation, address
RP	Name, date and municipality of birth, sex, address

Death certificate

Informants (two)*	Name, age, occupation, address or municipality, signature
* Often related to the RP	
RP	Occupation, civil status, address of decease, municipality of residence
Spouses of RP	Name, age, occupation, civil status, address or municipality of residence

Personal card (only for RP's still alive in 1940)

RP	Occupations, civil status (with dated changes), addresses and municipalities of residence, (dated), municipality of decease
Spouses RP	Name, date and municipality of birth
Parents RP	Municipality and date of birth and decease
Children RP	Name, date and municipality of birth, date and municipality of administrative removal due to death, marriage or departure

Marriage certificate

RP	Date and municipality of marriage, municipality of residence and address, occupation at marriage, names of previous spouses, signature
Illegitimate child RP	Names, date and municipality of birth
Spouse RP	Name, address, age, occupation, signature
Parents RP	Occupation, municipality of residence or date and municipality of death, signature
Parents in law of RP	Name, age, occupation, municipality of residence or date and municipality of death, signature
Witnesses (four)*	Name, age, occupation, municipality of residence, signature
* Often related to the RP	

Census registers (period 1810-1840)

RP	Occupation, civil status, address, religion
Parents RP	Name, age or date and municipality of birth, occupation, address/municipality of residence, etcetera (see RP)
Relatives RP (in particular siblings)	The same
Spouses RP	The same
Children RP	The same
Other persons*	The same

* Other persons living in the household, both related and unrelated (in particular servants)

Registers of conscripts

RP	Occupation at the time of examination, address, description, examination report (e.g. height), date of conscription
Father RP	Occupation, address

Capitations

RP	Assessment and estimated income, also important for evaluating occupational position
Father RP	The same

Land register

RP	Possession of real estate, assessment land tax
Father RP	The same

Church records

RP	Migrations (attestations), type of membership (only baptism or also confession)

Population register (1850-1940)

The population register always contains the following variables for all persons included:

- date of entry
- full name
- date and municipality of birth
- sex
- relation to the head of the household
- civil status (with possible changes and municipality where these changes have taken place)
- religion
- occupation
- date and municipality of provenance
- date and municipality of destination
- (possible) municipality of legal residence and further remarks

The outstanding feature of the population register is that it presents the RP in constantly changing stages in the life course. The following example gives a typical sequence:

1) as a son or daughter of the head of the household,
2) living independently or living with another household (for instance as a servant)
3) as a head of one's own household (or as a wife)
4) living as an elderly father or mother with the household of a child

The population register has a dynamical character. Frequent movers may have up to fifty entries in the registers during their life.

The HSN distinguishes between four types of data collection:

A First entry in parental household
B Itinerary (survey of all entries without actual storage of the information)
C Complete storage of all entries in the population registers of own household of RP
D Complete storage of all entries of the parental household (from birth RP onwards)