This is the published version of a chapter published in *Tradition and transition: Studies in microdemography and social change*.

Citation for the original published chapter:

The Demographic Data Base at Umeå University.

N.B. When citing this work, cite the original published chapter.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-132350
The Demographic Data Base at Umeå University

Christina Danell

As early as the seventeenth century Swedish ministers created an information system covering all individuals above six or seven years of age in their parishes, the so called 'examination registers'. By the middle of the eighteenth century these registers contained every one, including the small children, and they became the primary sources for population statistics delivered to the Swedish 'Table Commision'. As in the past, the population registers in Sweden are still kept by the clergy and used for different statistical purposes.

When historical demography, and especially 'microdemography' using family reconstitution, became a popular subject in many countries, it was decided that something had to be done in order to make the information in Swedish Church Books more easily available to researchers, especially since their information was unusually abundant. This was the reason for the establishment of the Demographic Data Base at the University of Umeå in July 1973.
The purpose of the Data Base

The Demographic Data Base deals with the collection, registration and computerising of data about individuals in the Swedish church records. The aim of the Demographic Data Base, according to the government regulations founding it, is:

- to register and transform demographic and social historical data for research, educational and archival purposes and to make these data available to researchers.

The Data Base should encourage scientific collaboration using the data and, where suitable, develop methodology.

The regulations are dated 1978-06-29 and apply from the first of July the same year. It was then the Data Base got its present permanent organisation as a special unit at the University of Umeå.

Organisation

The Data Base is led by a planning board with representatives from institutions connected with the work of the Data Base including: the universities, the archives, the museums, the Labour Market Board, the Central Bureau of Statistics and the personnel of the Data Base.

The Data Base is situated in Umeå and Haparanda. The management and the system unit are in Umeå. From the first of July 1980 there is also a research and development unit. The chair of Historical Demography will be connected to that unit. The professorship was created on the first of July 1980, but is not yet appointed. The production unit is in Haparanda.

Three Data Transferral Centers which are financed by the Labour Market Board for a limited time are also connected to the Data Base.

At the Data Base there are also two committees for co-operation and a reference group of five members of the planning board which deals more actively with the work.

In all there are at the moment about 60 employees at the Data Base, of which 17 work in Umeå and the rest in Haparanda.
The planning of the Data Base, the aims, the sources, the time period, the examination areas (parishes)

One of the principles for the planning of the Data Base was that it should be guided by the needs and wishes of researchers. Their opinions were collected in advance, among other things by an inquiry of possible users of the future Data Base. A great deal of proposals were collected. The aims of the data collection were decided by using these. These aims were roughly as follows:

The catechetical examination records and the other church records should form the base for the data collection. All information in the sources should be stored and to the greatest possible extent in plain language. The individual should form the smallest unit and each individual should be followed from one source to the other. Furthermore the individuals should be held together in families and households from year to year.

The nineteenth century was decided as the basic period for examination at an early stage. A limited number of parishes or groups of parishes (regions) were also to be chosen if the task was not to be too heavy. In the first stage seven parishes were chosen, widely scattered over Sweden. In the second stage 15 parishes around the city of Sundsvall were chosen. A third stage is now decided. It consists of the registration of about 35
parishes around the city of Linköping.

SOURCES

CATECHETICAL EXAMINATION RECORDS
RECORDS OF IN- AND OUT-MIGRATIONS
RECORDS OF BIRTHS AND BAPTISMS
RECORDS OF BANNS AND MARRIAGES
RECORDS OF DEATHS AND BURIALS


The work with the church parish registers

System design

When planning the Data Base the demands and needs of the researchers had a definite influence on the design of the system as well as on the registration process. The system design and the work with instructions and the registration process was done with the help of a research project at the Department of Education at the University of Umeå, 'Literacy and Education in Västernorrland 1750-1860', which was to examine, among other things, the church parish registers in the parish of Tuna in Medelpad.

The demands on a data base and a system for data collection were formulated in an article in Historisk Tidskrift 1973: 'Facts for research. The planning of a demographic data base.' as follows:

1. The system should be true to the sources. That is, it should be possible to compare and control the data collection by the authentic sources. And it should be possible to distinguish and analyse each separate source in the data processing.

2. The system should be complete. That is, all essential, relevant information should be included in the data collection.

3. The system should be open. That is, it should be possible to complete the data collection using longer time periods, within larger areas and with additional parallel sources. This means that it should be possible
to compare and co-ordinate data from different areas in an integrated information system.

4. The system should be economic, that is, economic in the manual registration process as well as in the system design and the automatic processing and storing.

5. The system should be directed to research, that is, designed to serve research in genealogy as well as in demography. This means that only listing of data or genalogical cataloguing of information is not enough. It should be possible to analyse the total population within each chosen area (parish) for a single year and for a certain time period.

The two first demands and the last directly concern the system design. They have been formulated as follows:

- Data should be collected completely and with high fidelity to the sources. It should be possible to reconstruct each source.
- A biography should be done for every person in the sources and also a family reconstitution.
- The system should be designed in such a way that it will be possible to do cross-sectional studies of the whole population as well as longitudinal analyses of demographic changes.

The two last demands refer to the demand of research-direction, that is to the requirement of the possibility of doing genealogical studies as well as demographic studies.

Demands of the system

<table>
<thead>
<tr>
<th>Archives</th>
<th>Reconstruction of the sources with a high fidelity and completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>The individual is to be described with biographies for longitudinal analyses and cross-sectional studies - for demographic research</td>
</tr>
<tr>
<td></td>
<td>Family reconstitution possibilities - for genealogical research</td>
</tr>
</tbody>
</table>
The process

Preparation of sources
Excerpt process

The information in the sources is written on cards
Linkage process

The cards are linked into 'individuals'

Data entry
File storage

Keying and storing of data

Retrieval

Retrieval of information

The result

Parish registers; catechetical examination registers etc

One card refers to one notation about one individual in one source

Biographies on a collection of cards from different sources

Sources and biographies on different files on discs

Information about sources and individuals on lists, tables or in other forms i.e. interactive applications
A short description of the process

When the decision is made that a certain parish is to be done by the Data Base the work starts at the Production Unit at Haparanda. Basically what is done is to transform the information in the church records to magnetic tape which is delivered to the System Unit at Umeå. There, the information is stored in files which make it possible for researchers and other interested people to order data-material, statistics, lists and so on.

THE PROCESS

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREPARATION</td>
<td>RECONSTRUCTION OF THE SOURCES</td>
</tr>
<tr>
<td>TRANSFERRAL</td>
<td>BIOGRAPHIES</td>
</tr>
<tr>
<td>LINKING</td>
<td>FAMILY RECONSTITUTION</td>
</tr>
<tr>
<td>DATA ENTRY</td>
<td></td>
</tr>
<tr>
<td>FILE STORAGE</td>
<td>STORING AND RETRIEVAL OF INFORMATION ABOUT</td>
</tr>
<tr>
<td>RETRIEVAL</td>
<td>SOURCES AND INDIVIDUALS</td>
</tr>
</tbody>
</table>

At first the sources are prepared, that is the sources are examined, checked and described. The result of this work is a description of the sources. This is done to make the interpretation of the sources more reliable and to make the transformation of the information in the sources fulfill the demand of fidelity to the sources at the end.

The transferral is the next thing done. That is, the information from the sources is written on cards to make possible keying and storing of data.

- One card corresponds to
- One notation
- on one line
- about one person
- in one source

When the transferral is done the cards are sorted and linked together in such a way that all cards that belong to one person lie physically together. A unique identity number for each individual is given. (This is a condition for the automatic processing of data). When the linking process is done all cards concerning one individual are linked together. The data is ready for data entry.

Data entry is the process of transferring data on cards to magnetic tape (keying). In this process simple transcription errors in the data are checked by using a computer. The output from the Production Unit is data which is linked to individuals and checked.

The System Unit is responsible for the further processing of the data, that is storing of the data to fulfill different demands for retrieval of information.

The data base system

The different demands of retrieval of data was the basis for the design of the data base system. The main demands - reconstruction of the sources with high fidelity and the possibility of demographic analyses - generate different demands for data structure and data files. The first demand means that the total collected data should be accessible as they are recorded in the sources and that it should be possible to identify each note about an individual. Because of the nature of the sources the total data file will include data which is superfluous in demographic analysis. The notes in the sources could also be contradictory, incomplete and illogical in this connection. To make it possible, under these conditions, to describe the reality behind the sources it is necessary to choose among the available data and make up new data when the sources are imperfect. The principles for choosing and construction of the data are not evident. In the system design of the Data Base this is done as follows. A standard version of constructed biographies is offered, leaving each investigator free to create biographies according to his own intentions on the basis of the source material. Thus the possibility of standardising the extracts from the Data Base means a reduction of costs for the investigator who can accept those principles of construction of biographies that have been
followed at the Data Base. The Data Base system that has thus been built up at the Demographic Data Base consists of two separate file systems: one source system and one event system. Besides this are also a number of special files, first of all files of codes and texts.

Source File System

In the source files there is a file for each type of source. In the source files are stored data with the individual as the unit. Each such unit comprises all data that have been found in the source in question. Every single note in the sources can be identified in the unit. The variables are stored exactly in the form they had when the excerpt was made; the sequence of the entries fixed at the compilation is maintained and forms the basis for creating biography. The link between the different source files consists of the identity given to each individual at the compilation. From considerations of space, annotations and causes of deaths are separated from other data and stored in special files for annotations and causes of deaths. The links between individuals which are established in connection with the compilation are stored in a Relations File.

Source Files

One file for every type of source
- Parish Catechetical Examination File
- Register of Births File
- Register of Deaths File
- Register of Marriages File
- Register of Migrations File

As well, there are four more sources of files, namely:
- Relations File
- Identity File
- Annotations File
- Causes of Deaths File

The main functions of the source files are
- To reconstruct the sources in their original form
- To make it possible to create biographies of the individuals to be found in the sources
To make possible family reconstitution by storing links between individuals

Event File System

The event file, which is a key file in the system, contains adapted data of individuals, made on the basis of the source files. The selection of variables which are included in a complete biography represents, beside identity data, demographic events and the status of the individual in different respects and at different points in time.

Biographies

A complete biography should include the following variables:

Identity information: Birth data, name(s), legitimate or illegitimate birth, twinship

Demographic events: Date of birth, death, marriage ceremony, change of residence within parish or across parish border, previous and later domicile

Status at given points of time: Trade (profession), social status, civil status, residence, position in family or household

Relations to other individuals: Parents, husband/wife, children

Other information: Volume/page/line in Parish Catechetical Examination Register, where the person has been inserted, reference to other sources, holy communion, catechetical examination, testimonials, vaccination

Adaptations made aim at creating a logical and reliable biography of each individual using the available material to the utmost. The principle is that, when double information is available, priority be given to information from the source of a certain type of events, supplemented by information from the Catechetical Examination Register. When the same information is found several times in the Catechetical Examination Register the first annotation is usually quoted. It should be possible to verify every entry in the Event File by a reference to the source. The Event File should
also contain links between individuals, to make possible reconstitution of families and genealogical research. Those links which are stored in the Relations File are checked and supplemented in connection with their being transferred to the Event File.

Retrieval

Storing of data, first of all Source Files and Event File, aim at meeting different requirements of retrieval of information. The Source Files make it possible to make extracts of variables (pieces of information) from the different sources, the Event File makes it possible to make extracts dealing with private individuals. There a is also a file for 'group extracts', i.e. extracts of data dealing with groups of individuals. Retrieving of data is done through batch processing and through inter-active questionnaires (i.e. questionnaires for direct communication with 'the computer' via terminal). Programs for standard extracts from the Data Base are rare. The inter-active programs aim mainly at giving the investigator an idea of the possibilities of the material. Most extracts are specially adapted to the desiderata of the investigators. The most frequent method of presenting data is still in the form of traditional data lists with listing of data referring to individuals or aggregated tables. It is also possible to order a selection of variables to be delivered on magnetic tape which the investigator himself has to supply. The possibilities of making individual inter-active selections of variables will increase in future. Other forms of presentation of data will also occur more frequently.

Time Schedule for Production

The first seven parishes

The first seven parishes are scattered all over Sweden from Skåne in the south to Norrbotten in the North:
Fleninge in Skåne
Locknevi in Småland
Gullholmen in Bohuslän
Trosa in Södermanland
Svinnegarn in Uppland
Tuna in Medelpad
Nedertorneå in Norrbotten
The work with these parishes is practically completed. What remains is documentation and certain quality checks before this research material can be fully made use of. Since the nature of the work with these parishes has been to some extent experimental the material has some deficiencies and limitations compared to that of parishes later dealt with.

The Sundsvall region

The parishes in the Sundsvall Region will be dealt with by the Production Unit according to the following schedule (except Tuna which is the first parish):

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lögdö</td>
<td>Selånger</td>
<td>Alnö</td>
<td>Ljustorp</td>
<td>Sundsvall</td>
</tr>
<tr>
<td>Tynderö</td>
<td>Njurunda</td>
<td>Timrå</td>
<td>Sättna</td>
<td></td>
</tr>
<tr>
<td>Lagfors</td>
<td></td>
<td>Timrå</td>
<td>Skönn</td>
<td></td>
</tr>
<tr>
<td>Attmar</td>
<td>Svartvik</td>
<td></td>
<td>Hässjö</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gullström</td>
<td>Indal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After the work of the Production Unit remains the work of the System Unit at Umeå to arrange the material in different files. This work is not yet planned. How long it will take and whether the order between the parishes will be maintained as indicated above is not yet decided. The whole region will be available towards the end of the eighties.

The Linköping region

The work with the Linköping Region is scheduled to begin towards the end of 1982 with transferring of birth registers. On the assumption that the present number of employees will be maintained, i.e. with the Transferral Center at Pajala assisting at this work, it is calculated to take about 11 years at the Production Unit.

Work with other source material

Beside the main work, that with Church archives material, there are also other series of sources available at the Data Base, e.g. statistics sent in to the rural deans for further delivery to the Cathedral Chapters and Central Government Departments (Table Commission 1749-1859 and Central Bureau of Statistics 1860 and onwards). These statistics are not published
at the local level, only at the County level. At the Demographic Data Base there has been built up, as an experiment, a special computer system to take care of and use this statistical material - Rural Deans' statistics, 'Table Commission Material'. The experiment comprises certain selected years and for these a selection of information has been registered for all rural dean districts in Sweden.

Similar statistical material at parish level is, as an experiment, scheduled to be registered in a special computer system - Parish Statistics, 'Central Bureau of Statistics Material'.

Another example of source material at the Data Base are official but unpublished School Statistics, 1874-1881, for all rural dean districts, counties and dioceses. These are to be found in special data files.

How to get material from the Data Base

The material of the Data Base is available to all investigators. To get the best possible result it is important to contact the Data Base at an early stage, if possible at the stage of the planning of the investigation. Within the Research and Development Unit a general strategy has been worked out in order to contact and advise new investigators. Experience has proved that in order to reach the best possible results research contact should be carried out by certain definite steps which are indicated below. A general strategy can of course not be followed strictly at each individual research contact, but it may serve as a description of how the co-operation of the Data Base with researchers and clients is supposed to work.

Strategy for Research Contacts

1. Mutual understanding of problem, formulation of problem, revision of research designs, selection of area, time groups, sources, variables etc

2. Introduction to previous research, litterature, the Data Base as research environment etc

3. Work with archives conditions, sources, microfilm, microfiche, basic statistics, parallel sources etc

4. Work with inter-active programs for standard extracts of information
about private individuals from Church Records and of statistics from Table Commission material, feed-back to sources

5. Work with simple programs for extracts and computer handling in connection with client's own files

6. More complicated extracts, using the System Unit

7. Discussion of results with the System Unit, evaluation of co-operation with the Data Base, feed-back of experience to the Data Base

Summary

About three years have passed since the Demographic Data Base became a permanent unit of Umeå University. At the beginning much of the time had to be spent to solve the special problems arising from the fact that a relief project had to be turned into a permanent institution. Many of the difficulties during the first years were due to the management, the research service and the System Unit being understaffed, since the character of the Data Base had been that of relief work. To the Data Base has been added a special Research and Development Unit and this means that the 'know-how' and the research amenities of the Data Base have been given a wider field of activities. This involves a concentration on research service, i.e. continuous co-operation with those researchers who make use of the Data Base. Another branch of this scheme is internal research work to illustrate the ability of the Data Base to make essential research contributions of different kinds.

A purposeful effort in the training sector is, indirectly, implied in the regulations and is also needed if the material of the Data Base is to find users in the future. A training program has been drawn up for this purpose.